ADA080835

EVALUATIONS OF OPERATIONAL DECISION AIDS

I. The Strike Timing Aid

Arthur I. Siegel
Edward G. Madden

Applied Psychological Services, Inc.
Science Center
Wayne, Pennsylvania

prepared for

Engineering Psychology Programs
Psychological Sciences Division
Office of Naval Research
Washington, D.C.

Contract N00014-77-C-0448
NR 198-018

January 1980

This work formed a part of the operational decision aid program of the Office of Naval Research.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

LEVEL II

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) (6) Evaluations of Operational Decision Aids. I. The Strike Timing Aid.		5. TYPE OF REPORT & PERIOD COVERED (9) Technical rept.
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) (10) Arthur I. Siegel, Edward G. Madden		8. CONTRACT OR GRANT NUMBER(s) NR 00014-77-C-0448
9. PERFORMING ORGANIZATION NAME AND ADDRESS Applied Psychological Services, Inc. Science Center Wayne, Pennsylvania 19087		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 198-018
11. CONTROLLING OFFICE NAME AND ADDRESS Engineering Psychology Programs Psychological Sciences Division Office of Naval Research Arlington, Virginia 22217		12. REPORT DATE (11) January ~~1980~~
		13. NUMBER OF PAGES 94 (12) 99
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

(15) N00014-77-C-0448

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Decision Aiding	Information Theory	Information Systems
Problem Solving	Management Decision Making	Command Support Systems
Information Processing		
Decision Making	Military Psychology	Information Analysis

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

As a part of the operational decision aid program of the Office of Naval Research, one of the decision aids (a strike timing aid) developed to date was tested to evaluate comparatively the merit, if any, of the full aid and its various components. The aid was developed to be representative of a class of operational decision aids which provide trend output information.

DD FORM 1473 1 JAN 73 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

402774

80 2 11 036

Flight "experienced" and flight "inexperienced" subjects solved "hard" and "easy" strike launch time problems using the full aid, selected portions of the aid, and without the aid. The results were analyzed by a variety of methods and supported contentions favoring the value of the aid. The results suggested: (1) an increase in decision validity by a factor of five, when unaided decisions were compared with aided decisions, (2) a quite strong achievement by the aid of its goals, and (3) differential effectiveness as a function of problem difficulty and the experience of the user.

Implications for aid development and evaluations of such aids are also presented.

EVALUATIONS OF OPERATIONAL DECISION AIDS

I. The Strike Timing Aid

Arthur I. Siegel
Edward G. Madden

Applied Psychological Services, Inc.
Science Center
Wayne, Pennsylvania

prepared for

Engineering Psychology Programs
Psychological Sciences Division
Office of Naval Research
Washington, D.C.

Contract N00014-77-C-0448
NR198-018

January 1980

This work formed a part of the operational decision aid program of the Office of Naval Research.

ABSTRACT

As a part of the operational decision aid program of the Office of Naval Research, one of the decision aids (a strike timing aid) developed to date was tested to evaluate comparatively the merit, if any, of the full aid and its various components. The aid was developed to be representative of a class of operational decision aids which provide trend output information.

Flight "experienced" and flight "inexperienced" subjects solved "hard" and "easy" strike launch time problems using the full aid, selected portions of the aid, and without the aid. The results were analyzed by a variety of methods and supported contentions favoring the value of the aid. The results suggested: (1) an increase in decision validity by a factor of five, when unaided decisions were compared with aided decisions, (2) a quite strong achievement by the aid of its goals, and (3) differential effectiveness as a function of problem difficulty and the experience of the user.

Implications for aid development and evaluations of such aids are also presented.

SUMMARY

As a part of the operational decision aid program of the Office of Naval Research, one of the decision aids developed to date, the strike timing aid, was subjected to a test to evaluate comparatively the merit, if any, of the full aid and its various components. This evaluation is one of a set of evaluations of various decision aids developed under the decision aid program of the Office of Naval Research.

Description of the Strike Timing Aid

The strike timing decision aid was developed by Analytics, Inc. to be representative of a class of operational decision aids which provide trend output. The version tested was not considered to be in a "ready for use" state. Essentially, the aid is based on a mathematical engagement model which predicts the outcome of an air strike as a function of strike launch time. On the basis of input information, the aid provides two types of user oriented information: (1) projected strike outcome information, and (2) expected strike utility. The projected outcome information consists of such items as projected own losses, projected enemy air losses, and projected enemy ground losses. The expected utility information presents the "value" of a strike as a function of strike launch time. This value is calculated as a function of "subjective" values assigned by the aid's user to the loss or destruction of various types of units and the number of units (both own and enemy) expected to be destroyed at various launch times.

Other user oriented features are also provided by the aid, e.g., an analysis of losses by mission segment and a sensitivity analytic feature.

Method

Flight "experienced" and flight "inexperienced" groups were asked to solve "hard" and "easy" strike launch problems using the full aid, selected portions of the aid, and no aid. The results were analyzed relative to five hypotheses concerning the utility of the aid. The hypotheses concerned: (1) the effectiveness of strike launch time decisions made with the use of the aid and those made without the use of the aid, (2) the perceived usefulness of the aid, (3) the effectiveness and perceived usefulness as a function of the experience of the user and as a function of problem difficulty, (4) the validity of the aid, and (5) the effectiveness of decisions made when only portions of the aid are made available for use.

Data pertinent to each of these five hypotheses were collected and compared with criterion data reflecting the optimum solution to each problem. Each participant in the study was also interviewed concerning his reaction to the various features of the aid.

Findings

The results supported contentions favoring the value of the aid. There were consistent, statistically significant differences, favoring aiding, between the unaided condition and some level of aiding. The data suggested an increase in decision validity by a factor of five, when unaided decisions were compared with aided decisions. The results of a multiattribute utility analysis indicated that the aid achieved its goals quite well.

A set of regression analyses indicated that the aid users did not employ all of the information provided by the aid when they attempted to solve a problem. More typically, the user selected one or two aspects which were important to him (e.g., weather at target, enemy air defense readiness, projected own losses) and based his final strike launch time choice on that (those) considerations.

There was some evidence of a differential effectiveness of the aid as a function of problem difficulty and experience of the user.

The interview information also provided support for contentions favoring the value of the aid. While some reservations were expressed about the form of certain output displays, most experienced participants indicated that they would use such an aid in an actual operational situation--at least as a supplement to other information on hand.

TABLE OF CONTENTS

TABLE OF CONTENTS (Cont.)

LIST OF TABLES

LIST OF EXHIBITS

LIST OF FIGURES

I. INTRODUCTION

Since 1974, the Office of Naval Research (ONR) has been investigating the feasibility of producing tactical level, computer based decision aids for application in various operational situations. From the outset, three important factors governed the program's development. First, the aids were intended to meet the needs of task force level decision makers and planners--task force commanders and their staffs. Second, new methodologies were emphasized which could handle the complexity inherent in naval command and control and in tactical planning. Third, the aids were to be objectively tested and evaluated.

Objective, experimental testing and evaluation were emphasized on at least three levels (Sinaiko, 1977). During the early stages, each aid was to be tested by the persons responsible for the aid's design and development in their own facilities. When an aid was considered sufficiently ready, it was to be evaluated by an independent agency. Finally, the appropriate naval user organization will test any aid being considered for use in the Fleet.

Organization of Decision Aid Program

The essential structure of ONR's operational decision aid program rests on the activities of a variety of university and industrial research and development organizations. The total program is monitored by ONR through a steering committee. The various participating organizations are primarily involved in the development of the aids although some have also addressed specific problems that bear on the general nature and effectiveness of such aids (e.g., Lucas and Ruff, 1977; Analytics, Inc., 1976; Brown, 1978).

The Department of Decision Sciences, Wharton School, University of Pennsylvania supplies computer and data management support systems as well as facilities and apparatus for demonstrating and testing the aids.

Applied Psychological Services is the organization selected to evaluate the aids independently. The role of Applied Psychological Services is to serve the function of a crucible--to test critically, rigorously, and fairly each of the decision aids and to report findings and recommendations for improvement of the various aids. Such evaluations are to be conducted within the context of the ONRODA scenario (Payne and Rowney, 1975; Rowney, 1975), and with naval personnel serving as test subjects.

Aid Evaluation

Because it seemed important to place the present evaluation program into the context of a total decision aid developmental framework, a conception (Figure 1) of the steps to be followed during the development of such aids was developed (Siegel and Madden, 1979). The figure is read from the bottom to top with the considerations involved in each stage entering from the left of each bo.. and the results of each stage exiting to the right. The number(s) above each Figure 1 box represent criteria which may be applied after each developmental stage. These criteria are defined in Table 1. The rounded boxes associated with each rectangular, stage box represent descriptors which may be applied as the criteria at the successive stages are met. Accordingly, an aid may be successively called "suitable," "testable," "reasonable," "valid," "effective," and "useful." Note that we are primarily concerned within the present aid evaluation program with the upper right box--"validation testing-exercise in lab experiment and compare with intermediate criterion."

Within this context, the work possesses a number of characteristics:

- use of well-controlled, precise, multivariate methods
- programmatic approach
- full coordination with ONR
- orientation towards possible conditions of actual aid use in the Navy
- coordination with aid developers but maintenance of evaluation integrity
- use of previous developed ONRODA action scenarios where possible

Purpose of Present Work

The global purpose of the aid evaluations is to answer for each aid such questions as:

- Does it work?
- Why does it work?
- How can it be made to work better?

It is important to know not only that an aid does or does not possess utility but also which of its characteristics contribute to the utility. For example, an aid might be useful because it synthesizes information from a

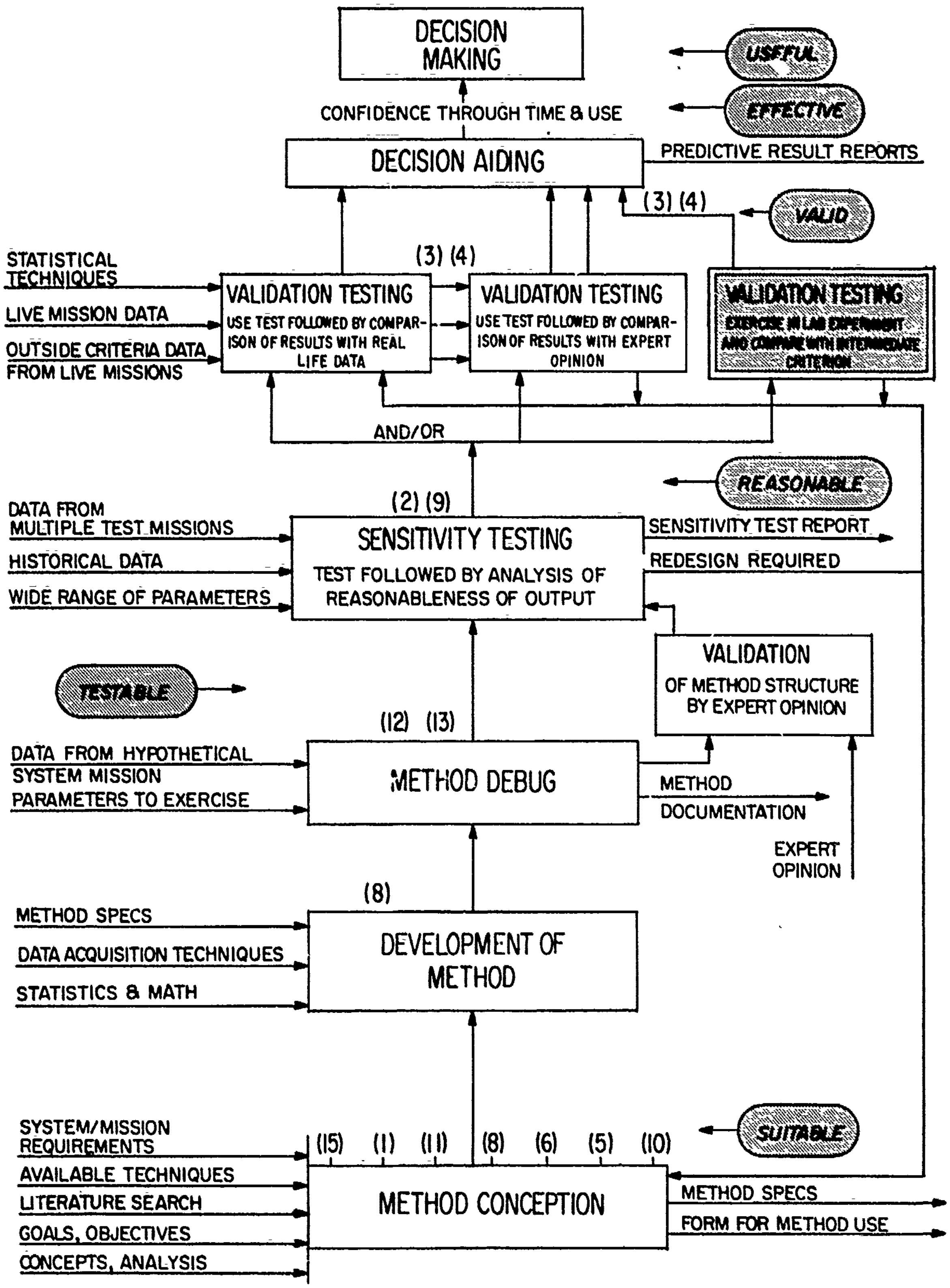

Figure 1. Sequence of aid development.

Table 1

Criteria for Evaluating the Utility of a Decision Aid

Criterion	Definition
1. Internal consistency	Extent to which the constructs of the aid are marked by coherence and similarity of treatment
2. Indifference to trivial aggregation	Potential of the aid to avoid major changes in output when input groupings or conditions undergo insignificant fluctuations
3. Correct prediction in the extreme (predictive or empirical validity)	Extent of agreement (correctness of predictions) between the aid and actual performance at very high/low values of conditions
4. Correct prediction in mid range (predictive or empirical validity)	Like above for middle range values of conditions
5. Construct validity	Theoretic adequacy of the aid's constructs
6. Content (variable parameter) validity (Fidelity)	Extent to which the aid's variables/parameters match real life conditions
7. Realism or "face validity"	Extent to which selected content matches each attribute included
8. Richness of output	Number and type of output variables and forms of presentation
9. Ease of use	Extent to which an analyst can readily prepare data for, apply, and extract understandable results from the aid
10. Cost of development	Value of effort to conceive, develop, test, document, and support
11. Transportability-generality	Extent of applicability to different systems, missions, and configurations
12. Cost of use	Value of all effort involving use of aid including data gathering, input, data processing, and analysis of results
13. Internal validity	Extent to which outputs are repeatable when inputs are unchanged
14. Event or time series validity	Extent to which aid predicts event and event patterns

diversity of areas to give a planner new insights; or, an aid might significantly reduce the amount of time and labor required to make a decision; or, it might allow for a more careful analysis of a broad range of alternatives. Therefore, in attempting to examine the usefulness of any aid, it is necessary to specify the factors out of which its utility may have been derived and how they may interact.

Linked with usefulness is a concern for the goals of an aid, their relative importance, and how closely they were achieved. The goals are objective expressions of what the aid should be, or do, or facilitate. Therefore, an examination of how clearly the goals of an aid were achieved and their relative importance should lead to a better understanding of what contributes to the usefulness of an aid and poss.bly to how to improve it.

Other Evaluative Considerations

Other aspects of a decision aid which must be considered in any thorough evaluation are of a less general nature than those already discussed. But, they make important contributions to the assessment of an aid. These considerations relate very strongly to human factors considerations and include three general groups:

(a) Nature of information

(1) Is sufficient information provided by the aid?

(2) Is the information provided pertinent? Accurate? Timely? In the required form?

(b) Method of output presentation

(1) Is the information presented in a manner which is responsive to user requirements?

(2) Are the tables/graphs and other output format easily comprehensible?

(3) Is an optimal amount of information presented in each table/graph?

(c) User ease

(1) Is the aid relatively easy to use?

(2) Are the user commands to the computer system arranged so as to minimize both effort and confusion?

(3) Are the user oriented error messages understandable and informative?

The Strike Timing Decision Aid

This report presents the methods, procedures, and results of the first in a series of decision aid evaluative studies. The report is concerned with an evaluation of Analytic's Strike Timing Decision Aid (ASTDA). The ASTDA was designed as a tactical decision aiding system to be used by task force level flight operations officers. Essentially, the ASTDA is based on a mathematical engagement model which predicts the outcome of Blue (own) air strikes launched against Orange (enemy) forces. The aid was developed by Analytics, Inc., within the framework of the general ONRODA scenario (Payne and Rowney, 1975).

ASTDA Characteristics

ASTDA was developed as a prototype of a class of aids which might support any decision regarding when to take an action when the action itself has already been determined. In its specific implementation, ASTDA is intended to supply flight operations officers with information concerning: (1) likely combat conditions (e.g., weather at target, number and type of Orange forces, etc.) at various future points in time, (2) probable outcomes (e.g., number and kind of Blue forces lost, etc.), and (3) expected utility (the relative value of an air strike to Blue) of air strikes launched at various future points in time.

Several other sets of information are also made available by the ASTDA. These allow the operations officer to examine more finely other aspects of the outcomes of a projected air strike. One of these is losses by mission segment. This information indicates the Blue losses from a specific air strike as a function of air mission segment (take-off, ingress, at-target, egress, and landing). The other is a sensitivity analysis which allows an examination of the sensitivity of the expected utility to changes in Blue or Orange forces.

All the information available from ASTDA can be presented in both the tabular and the graphic forms. The information displays, except those pertaining to weather conditions which are presented as probabilities, are presented as means and delta biased uncertainty bands. A delta biased uncertainty band is a concept, developed by Analytics, which shows the two standard deviation interval around the mean "whose midpoint has been moved away from the . . . 'mean' of the distribution by an amount given by the parameter delta" (Glenn, 1978). The value of delta used is selected with the purpose of correcting for skewness. Accordingly, a delta biased uncertainty band never includes values that are not actually within the range of the distribution.

ASTDA Information Base

The ASTDA processes relevant information to calculate air strike result predictions for various strike launch times. This information would normally be supplied by weather officers, readiness officers, intelligence officers, etc.

The ASTDA requires five categories of input (data base) information: the first three concern the strength and resources of the Blue and the Orange forces, while the last two concern weather conditions. The information for the first three categories are entered as means and standard deviations. ASTDA converts the standard deviations into delta biased uncertainty bands. The information concerning the weather is entered in terms of the probability of good visibility. Once the information is entered, it can be called and displayed in either the tabular or the graphic forms. The displays which are then available are the Blue Force Availability (BFA), the Orange Air Defenses (ORAD), the Orange Ground Defenses (ORGD), the Weather at the Target (WAT), and the Weather at the Carrier (WAC) as a function of time. For the purposes of the present evaluation, the input information was supplied to the subjects and was preinserted into the system.

In the present work, the BFA displays indicated the relevant information for one type of Blue fighter-interceptor (the BF1s), and for two types of attack-bombers (the BB1s and the BB2s). The BFA displays also included information on the desired number of Blue aircraft (DNB). The ORAD displays contained the information concerning two types of Orange fighter-interceptor (the OF1s, and the OF2s). The ORGD displays included information on two types of ground defense: OD1 (surface-to-air missiles), and OD2 (anti-aircraft artillery), as well as the number of OT1s (passive ground targets). The WAT and WAC displays indicated the probability of good visibility at the target and at the carrier, respectively. The probability of good visibility at the target was shown graphically as a function of time at the target while that for visibility at the carrier was plotted as a function of the landing time.

ASTDA Output

ASTDA produces two primary types of user oriented information: (1) projected outcome information, and (2) expected strike utility. As stated previously, this information is presented as a function of various strike launch times and is made available in both the graphic and the tabular formats.

a. Outcome. The information entered into ASTDA is used in an engagement model which predicts the results of air strikes launched at the various strike launch times. The output of the engagement model is essentially a statement

of the probable number and kind of lost or destroyed Blue and Orange forces. The strike outcomes in the evaluation were available in three displays: Blue Force Losses (BFL), Orange Air Losses (ORAL) and Orange Ground Losses (ORGL). The likely number of lost BF1s (fighter-interceptors), BB1s and BB2s (attack-bombers) across each prospective strike launch time was presented in the BFL displays. The ORAL displays indicated the number of OF1s and OF2s (fighter-interceptors) likely to be destroyed at each prospective strike launch time. The ORGL displays contained the number of OD1 (surface-to-air missiles), OD2 (anti-aircraft-artillery), and OT1 (passive ground targets) which would probably be destroyed at each of the strike launch times. The information contained in each of these displays was given as both means and delta biased uncertainty bands.

b. Expected Utility. In normal ASTDA employment, to determine the expected utilities of strikes launched at various times, the operations officer would first assign "subjective" values to the loss and destruction of each unit type. The value assigned to each unit would reflect the user's judgments of importance of the unit to the overall mission. In the present evaluation, these judgments were assigned independently of the subjects.

ASTDA computes the expected utility (EU) by taking the number of units of each type (Blue and Orange) predicted by the engagement model to be lost or destroyed, multiplying each by its assigned value, and summing across the units. The resulting sum is the expected utility and the higher the expected utility the better the relative outcome of the air strike for Blue. The aid computes an expected utility independently for each anticipated strike launch time. The calculated utilities are presented as a display in which utility is plotted as a function of the strike launch time. The strike launch time with the highest expected utility would therefore be considered the best time to launch an air strike.

Other ASTDA features, of a less direct nature, were not included in the present evaluation. These included the analysis of the losses by mission segment and the sensitivity analytic features.

Specific Purposes of Present Study

The evaluation plan employed for the ASTDA evaluation represents a synthesis of an initial plan developed by Applied Psychological Services and one proposed by Analytics, Inc. To reach a consensus, a series of

meetings was held. Representatives of Applied Psychological Services, Analytics, Inc., the Department of Decision Sciences of the Wharton School at the University of Pennsylvania, and the ONR participated in these meetings. A draft of the consensus was prepared (Siegel, 1978) and submitted to the other cooperating agencies. The final plan emerged as the result of comments by the other organizations on the draft.

The evaluative research plan was developed to test five hypotheses concerning the effectiveness of the strike timing decision aid under laboratory test. The five hypotheses were:

<u>Hypothesis 1.</u> More effective strike timing decisions can be made using the ASTDA than without the aid.

<u>Hypothesis 2.</u> Users will perceive the aid to possess value.

<u>Hypothesis 3.</u> Effectiveness and perceived value will not vary as a function of the user's operational experience level or problem difficulty.

<u>Hypothesis 4.</u> The strike timing decision aid possesses criterion related validity.

<u>Hypothesis 5.</u> Decision effectiveness will vary systematically as the characteristics of the aid are varied.

II. METHODS AND PROCEDURES

A five by two by two factorial design formed the basis for the present evaluation. The three independent variables were aid level, test problem difficulty ("easy" or "hard"), and Navy operational experience of the subjects ("minimum" or "considerable").

Each subject was presented with a series of scenario problems for which he was required to rank order launch times for an air strike from a carrier against the ONRODA island.

The main dependent variable was the launch time rankings of the subjects. The choice behavior was compared with criteria data to examine consistency with the expert Naval opinion and the best alternative indicated by utility values as indicated by the aid. The details are described below.

Factors

The first evaluation design factor was concerned with the levels of the decision aid. Five levels were investigated in an attempt to determine the effects of parts of the aid (aid features) in isolation as compared with the aid as a whole and with a no-aid (control) condition. The first level was the full-aid condition in which all five input (BFR, ORAD, ORGD, WAT, and WAC) and all four output displays (BFL, ORAL, ORGL, and EU) were made available to the subject. The second level was a utility condition in which the five input displays, (BFR, ORAD, ORGF, WAT, and WAC) and only one output display (EU) were made available. The third level was the outcome condition in which the five input (BFR, ORAD, ORGL, WAT, and WAC) and three outcome displays (BFL, ORAL, and ORGL) but not EU were available. The fourth was a no uncertainty bands condition in which the five input (BFR, ORAD, ORGF, WAT, and WAC) and four output (BFL, ORAL, ORGL, and EU) displays were presented but only as means, i.e., the delta biased uncertainty bands were deleted from the displays. The final level was an unaided control condition in which the subjects received only the input information (BFR, ORAD, ORGF, WAT, and WAC). This information was made available only in the tabular form for the fifth condition to simulate what might currently be available in the fleet.

Hypothesis 1 was tested by comparing the adequacy of subjects' strike launch time choices in the fully aided condition with the adequacy of their choices in the unaided condition. The three intermediate aid levels were included to test Hypothesis 5 and to isolate the contributions of the outcomes, the utilities, and the uncertainty bands to the decisions made in the full aid condition.

The second factor, problem difficulty, was varied over two levels. Each problem was classified as either "easy" or "hard." This served to test partially Hypothesis 3. The method of classifying a problem as "easy" or "hard" is discussed later in the section on "Problem Selection."

The final factor was included to allow at least partial test of Hypothesis 3--effectiveness and perceived value of the ASTDA will not vary as a function of the subjects' operational experience. To facilitate the testing of this hypothesis subjects were sampled from two populations: those with Naval flight oriented experience and those who have not had such experience.

Criterion Data

Within a test such as that described here, the criterion (standard against which the merit of the aid may be judged) choice represents a particularly difficult problem. Any criterion must possess such attributes as reliability, analyzability, objectivity, quantifiability, and acceptability. The first four of these criterion requisites are psychometric in nature and are technically manageable. The last requisite, acceptability, refers to the degree that others will accept the criterion as an index of merit or, alternatively, its relevance. Here value judgments come into play. At the extreme, the merit of the aid during wartime use might be the only acceptable criterion. Such a criterion is, of course, quite impractical. As one successively backs off from this ultimate criterion, he becomes more and more involved with intermediate criteria. Merit during a fleet exercise might represent on intermediate criterion that is quite proximal to the ultimate.

When one is involved with a laboratory test, as in the present work, the available criteria are more remote. Moreover, the conditions of a laboratory test, no matter how realistically they may simulate actual conditions, will only remotely resemble shipboard conditions and wartime stresses. The reader may ask, "What confidence may we have in such intermediate criteria?" The answer to this question seems to be that if the aid is shown to possess merit relative to the intermediate criteria, it may possess merit relative to more ultimate criteria. If the aid fails to possess merit relative to intermediate criteria, it probably will not possess merit relative to an ultimate criterion.

Two sets of standards or intermediate criteria were selected against which the strike launch time decisions of the subjects were judged. The first criterion was the launch time utility as predicted by the ASTDA. The second criterion was the launch time preferences as judged by a panel of experts. With these criterion data on hand, the agreement between the expert opinion and the utilities predicted by the aid could be examined and used as a measure of aid validity (Hypothesis 4).

Expert Opinion Criterion

The expert based criterion data were obtained through the c ra- tion of the ONR. Four Naval officers--two Captains, one Commander, and one Lieutenant Commander--volunteered to form a panel. The panel may be considered to be "experts" in that all members were senior in rank, sessed considerable operational experience, and were familiar with the nature of the strike timing problem and the ONRODA scenario. The panel met over a 1.5 day period. At the outset, the panel was briefed on the purpose of the aid evaluation, the ASTDA, the ONRODA scenario, assumed own and enemy force strengths and fighting capabilities and characteristics, weather conditions, and the problems inherent in selecting a launch time. Following this, each member of the panel was asked to work independently through 24 scenario problems and to indicate, for each, a launch time ranking and the difficulty of the problem.

After completing this independent work, the participants were asked to assemble as a panel and each scenario problem was reviewed. In the review, each participant indicated his launch time ranking, explained the basis for his decisions, and how difficult it was to make the choice. If one or more judges differed in the preferred launch time ranking for a given scenario, the panel attempted to understand why the difference occured. After this discussion, the participants were given the option to change their launch time selections. This was done in an attempt to obtain convergence and a consensus as to the best launch time ranking for each problem. In the case of lack of full convergence, the median launch time ranking of the experts for each scenario problem was subsequently used as the preferred launch time.

Aid Generated Utility Criterion

As a part of its internal logic, the ASTDA generates an expected utility. The aid generated utility values were employed as the second criterion in the present work.

Dependent Variables

As noted earlier, the primary dependent variable was the subjects' choices of preferred air strike launch times. These were compared with the criterion data.

Two other dependent measures were obtained: (1) a statement of the perceived difficulty of each problem, and (2) a statement of each subject's confidence in the correctness of his launch time ranking for each problem.

Exhibit I summarizes the evaluation design. Exhibit II summarizes the information made available in each of the five aid levels.

Exhibit I

Summary of Evaluation Design

Background	Difficulty	Aid Levels				
		Full Aid	Utility	Outcome	No Uncertainty	No Aid
Experienced	Easy Hard	n = 6	n = 6	n = 6	n = 6	n = 6
Inexperienced	Easy Hard	n = 6	n = 6	n = 6	n = 6	n = 6

Exhibit II

Information Made Available in Each Aid Level

	Information Provided			
Level	Input	Utility	Outcome	Uncertainty Bands
Full Aid	✓	✓	✓	✓
Utility	✓	✓		✓
Outcome	✓		✓	✓
No Uncertainty	✓	✓	✓	
No Aid	✓			

Subjects

Sixty subjects participated in the study. Half (30) of the subjects possessed Naval flight oriented experience and half (30) possessed no such experience.* The experienced subjects were recruited from a variety of sources, including Naval Air Reserve units, the Naval Air Development Center, and advertisements in local newspapers. Their Navy ranks were: Ensign = 1; Lieutenant = 12; Lieutenant Commander = 8, Commander = 6; Captain = 1; Marine Corps Captain = 1; and "unavailable" = 1.** The sample possessed a mean of 9.45 years in aviation, a mean of 2440 flight hours, a mean of 88 carrier landings, a mean of 10.5 years in the Navy, and a mean of 14 months of carrier duty.

The inexperienced subjects were exclusively midshipmen in the NROTC program at the University of Pennsylvania.

Experienced subjects, except officers on active duty, were paid $30.00, and inexperienced subjects were paid $10.00 for participating in the study.

Apparatus

The evaluation was conducted in the decision aid facility established by the ONR at the Department of Decision Sciences, Wharton School, University of Pennsylvania. The timing, presentation, and storage of experimental events was controlled by a PDP-10 computer. The subjects could enter commands into the system through a Data Media terminal and the resultant displays were projected on two screens. Tabular displays were presented on one Data Media screen and the color graphics, controlled by a Grinnell LSI-11 microprocesser, were shown on a parallel screen. Figure 2 presents the general equipment arrangement. As indicated in Figure 2, stations for system support personnel and the evaluation administrators were separated from the evaluation area. The evaluation conductors were able to observe the information displayed for the subjects and their activities by way of a special monitor.

Subject Orientation and Training

To provide a full, but standardized, orientation to each subject, a set of video tapes was prepared. Each subject, depending on the evaluation

*Although useful to strike planning, flight experience is not a critical prerequisite for strike planning.

**Where a subject was no longer on active duty, his rank on discharge is reported.

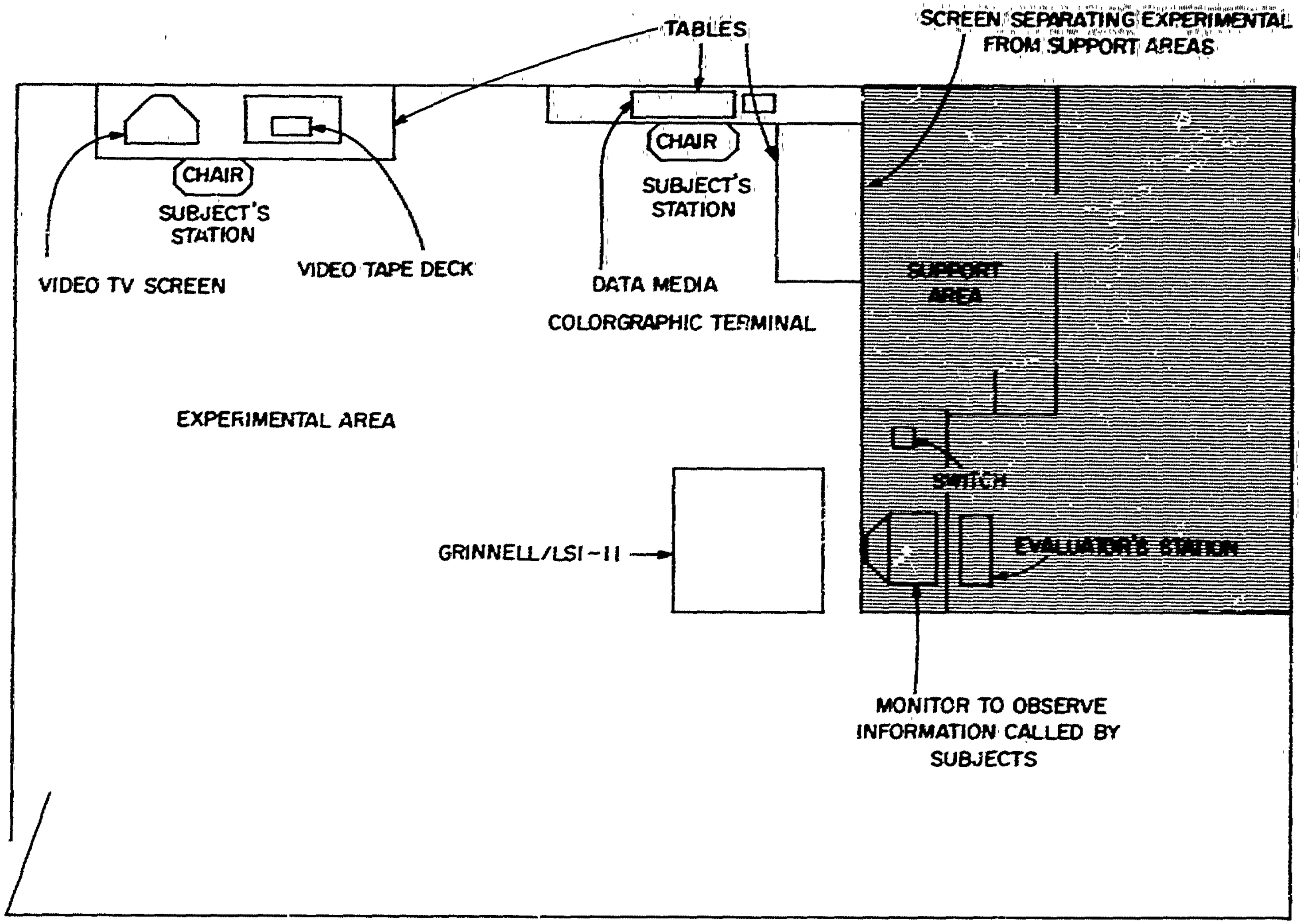

Figure 2. Schematic of area and equipment used in ASTDA evaluation.

condition to which he was assigned, was shown two tapes. One of the tapes provided a general discussion of the factors that the ASTDA considered to predict the outcome of air strikes, e.g., engagement characteristics and probabilities. All subjects viewed this tape. The use of the other tapes was restricted--each was oriented for the subjects assigned to a specific condition. In order to make the instructions received by each subject as standardized as possible, special care was taken with the construction of the tapes. The tape for the first condition, the full-aid condition, was used as a master tape. Wherever possible, the tapes for the other conditions were copies of the first except that, depending on the condition, certain scenes were edited out. For example, the scenes pertaining to the expected utility were deleted from the tape used for the subjects in the outcome condition level. Only rarely did the editing prove to be untenable in which case appropriate scenes were added.

Problem Selection

Analytics, Inc. developed the bank of 24 ASTDA oriented problems which were evaluated by the experts. Eight of these problems were selected for inclusion in the formal evaluation.

In order to satisfy the needs of the design, problem scaling was required along a difficulty continuum. The expert panel indicated that scenarios were easier which contained large differences among the consequences. Therefore, one basis for estimating the difficulty differences between problems was in terms of the spread of the possible outcomes across the strike times--the greater the spread of consequences the easier the problem. The alternative, that close solutions might be easier because a guess at a solution or tossing a coin to derive a solution would not cause serious differential consequences was not considered by the panel--probably because serious questions are not answered in a trivial way.

The aid supplies a general measure of the consequences of launching an air strike at any particular time as the expected utility. A statement of the differences between the expected utilities across strike launch times could accordingly be used as a gauge for specifying difficulty which would appear to be congruent with the guidance of the expert panel. Therefore, difficulty was relatively specified by the standard deviation of the expected utilities across the strike times in each problem in the bank. Four problems were selected which had utility standard deviations ranging from 2.81 to 4.06 ("hard" group) and four were assigned to the "easy" group which possessed standard deviations ranging from 7.94 to 13.85.

Exhibit III presents an example of one problem as it was presented to the subjects.

Exhibit III

Example of Strike Launch Time Problem

Strike Timing Problem 6

Your task is to select a strike launch time between 0600 and 1100 hours tomorrow for a cyclical strike against the Orange Forces on Onroda Island.

Blue Force Readiness

Unit Type	Number Definitely Ready At 0600	Number In Repair Which Might Be Ready By 1100	Average Number Expected To Be Ready At Each Time 0600	0700	0800	0900	1000	1100
BF1	10	2	10.9	10.9	10.9	10.9	10.9	10.9
BB1	5	1	5.6	5.6	5.6	5.6	5.6	5.6
BB2	13	3	14.9	14.9	14.9	14.9	14.9	14.9

Orange Air Defenses

Unit Type	Time of Encounter 0600	0700	0800	0900	1000	1100
OF1	10.7(7.8, 13.2)	9.8(6.6, 13.0)	8.7(5.4, 12.0)	8.0(4.8, 11.2)	7.5(4.5, 12.5)	7.3(4.4, 10.2)
OF2	21.9(17.3, 26.5)	19.7(13.7, 25.7)	17.8(11.7, 23.9)	16.4(10.6, 22.2)	15.4(10.0, 20.8)	14.6(9.7, 19.5)

Orange Ground Forces

Unit Type	Time of Encounter 0700	0800	0900	1000	1100	1200
OD1	8.4(6.6, 10.2)	8.4(6.6, 10.2)	8.4(6.6, 10.2)	8.4(6.6, 10.2)	8.4(6.6, 10.2)	8.4(6.6, 10.2)
OD2	10.7(8.7, 12.7)	11.1(9.1, 13.1)	11.5(9.3, 13.7)	11.9(9.7, 14.1)	12.2(9.9, 15.5)	12.6(10.3, 14.9)
OT1	20.7(17.3, 24.1)	20.7(17.3, 24.1)	20.7(17.3, 24.1)	20.7(17.3, 24.1)	20.7(17.3, 24.1)	20.7(17.3, 24.1)

Weather Conditions

Location	Time For Weather Prediction 0700	0800	0900	1000	1100	1200	1300
Target	.90	.87	.84	.81	.78	.75	
Carrier		.90	.87	.84	.81	.78	.75

Decisions, Judgments, and Ratings

A special computer routine was developed that was called by a subject when he was ready to record his decision(s) for a problem. The routine queried the subject about his choices by projecting questions on one of the display screens. The subject was required to rank order six potential strike launch times from the "best" to "worst." The first question asked the subject to indicate (from a choice of six) the time he thought best to alunch an air strike for the scenario involved in the problem. After the question was displayed, the system waited for an answer to be entered. Once the information was entered, the system asked for a second best strike launch time to be entered. After the strike timing decisions were collected, the routine then successively presented three more questions. The first question inquired into the subject's confidence in his decisions while the other two were concerned with the subject's perceived difficulty in reaching a decision for the problem. The first asked the subject to indicate his degree of confidence in his ranking along a five category confidence scale which was projected on one of the display screens along with the question. The scale ranged from "1" (not at all confident) through "5" (completely confident). The subject answered by entering, by means of the keyboard, the number appropriate to his level of confidence.

The first difficulty oriented question asked the subject to indicate perceived problem difficulty on a five point difficulty scale. The scale ranged from "1" (not at all difficult) through "5" (very difficult) and again the subject responded by typing in his judgment. The second perceived difficulty oriented question (third question) was structured by a magnitude estimation technique and required the subject to rate the difficulty of the problem he had just finished in relation to a modulus. The second practice (see "Procedure" section) problem was used as the modulus and assigned an arbitrary value of 100. It was expected that a scenario problem which was perceived to be twice as difficult as the modulus would be assigned a value of 200 while one perceived to be half as difficult would be assigned a value of 50. The subject indicated his rating by entering through the keyboard a representative number in response to the routine's query.

Procedure--Overview

Each subject was classified as either "inexperienced" or "experienced" and randomly assigned to one of the five aid levels.

The sequence of presentation of the scenario problems was randomized across subjects with the limitation that no two problems from the same difficulty category could be followed by a third.

Two data collection sessions, which lasted about three hours each, were required for each subject. The time was divided between an instructional and a testing phase. The instructional phase required about one hour.

During this time, two video tapes were shown and the subject used the aid to solve two practice problems. The testing phase followed. Following problem completion an interview inquiry was completed.

Subject Training

Prior to actual practice using the aid, each subject viewed two video tapes. The first tape was general and appropriate to all conditions. The second video tape shown to each subject was the one which was specifically fitted to the aid condition to which the subject was assigned.

The first tape initially discussed the purposes of the testing, what was expected of the subjects, the ONRODA scenario and the factors which the ASTDA considers in evaluating the outcome of an air strike. The factors included survival probabilities, engagement characteristics, and tactical considerations. The survival probabilities discussed were: (1) Blue aircraft survival probabilities against Orange air and ground forces in one-on-one engagements, (2) the conditional probabilities of Blue bombers evading Orange attack fighters during ingress and successfully returning to the carrier, (3) Orange air and ground force survival probabilities against Blue aircraft in one-on-one engagements, (4) survival probabilities for various numbers of Blue fighter/interceptors against various numbers of Orange attack fighters, and (5) survival probabilities for various numbers of Orange attack fighters against various numbers of Blue fighter/interceptors.

Other factors discussed in this tape concerned limitations on the number of simultaneous attacks of each unit and the tactical assignment of Blue units to various Orange ground targets. The survival probabilities, engagement characteristics, and tactical assignment were discussed for both "good" and "bad" weather contingencies.

One other factor which was described was the relative value to Blue of each lost or destroyed unit. These values are shown in Table 2 and were used by the aid to calculate EU. They were discussed to give the subject some concept of the relative value weighting of various units. In any operational version of such an aid, the user would be able to insert his own values for the various units, depending on the tactical situation, his own experience, and other factors.

The second tape made available information that was appropriate to each condition. The input information, BFA (Blue Force Availability), ORAD (Orange Air Defense), ORGF (Orange Ground Forces), WAT (Weather at Target), and WAC (Weather at Carrier) were described and demonstrated (this information was confined to the tabular form for the no-aid condition). The output information was specific to conditions. For the

full-aid condition, all the output information was described and demonstrated: BFL (Blue Force Losses), ORAL (Orange Air Losses), ORGL (Orange Ground Losses), and EU (Expected Utility). These were described and illustrated both as means and delta biased uncertainty bands in both the tabular and the graphic forms. In the expected utility condition, only the information pertinent to the EU and the related displays was presented. Likewise, in the outcome condition, the BFL, the ORGL, and the ORAL display were described and illustrated. Of course, in the no aid condition the output information was not even mentioned. For the no uncertainty bands condition, virtually the same information was discussed as for the full-aid condition except that no references were made to the measures of uncertainty normally given by the aid. Normally, in a tabular display, the ASTDA presents the mean and the delta biased uncertainty for each strike launch time, e.g., in the BFL display, among other things, data for the lost BF1s at a particular launch time might be presented as 5 (2, 7). The first number of the expression represents the mean and the numbers in the parenthesis are the lower and upper bounds of the delta biased uncertainty measure. In the no uncertainty condition, the same data were displayed as 5 (0, 0). To explain this ambiguity the relevant tape made a vague reference to a lack of variability. In the normal graphic presentations, the means were represented as points and the uncertainty bands were represented as bars extending above and below the points. Only the means (points) were shown and discussed in the no uncertainty condition.

Table 2

Equivalent Unit Value to Blue of the Destruction of a Single Force Unit of Each Type

Unit Type	Value
BF1	-6
BB1	-3
BB2	-1
OF1	1
OF2	4
OD1	1
OD2	1
OT1	5

Practice Session and Menus

After the two video tapes were viewed and questions about procedure, if any, answered, the subjects worked through two practice scenarios with the help of the evaluation administrator. The only difference between the practice scenario administration and the test scenarios--aside from the participation of the administrator--was that in the practice sessions the subjects were not required to make judgments of confidence and difficulty. The practice sessions served to familiarize further the subjects with the aid, the use of the equipment system, the various displays, and their task during the formal data acquisition.

The evaluation administrator acted as the system operator for the first practice problem. He demonstrated how to call displays by entering the appropriate commands. The commands were single, three to six letter words and were organized into two lists or menus--an input and an output menu. One list was available at a time and displayed at the bottom of one display screen. At the beginning of each problem, the input menu was available. It contained each command used to call up the input displays, i.e., BFA, ORAD, ORGF, WAT, and WAC. In addition, the list contained four other commands. The first, a HELP command, produced a list of all the available displays and the commands used to call them. The second, a DECIDE command, called the special routine which allowed the subject to record his strike timing decision(s), and to record his estimates of difficulty and of confidence. The third, a RUN command (not available in the no-aid condition), removed the input list and produced the output list. Finally, a RETURN command removed the output menu and restored the input menu.

The output menus' contents varied across conditions. The output menu contained commands for all the output data in the full aid and the no uncertainty conditions, i.e., BFL, ORAL, ORGL, and EU. In the outcome condition, the menu did not contain the EU command while in the expected utility condition the menu did not contain the BFL, ORAL, and ORGL commands.

After the first practice scenario was finished, the fact that the subject would normally be questioned at the end of each test problem on his confidence and the perceived difficulty of the problem was explained. A clear and concise description of the magnitude estimation technique was given and the use of the second problem as a modulus was discussed. Any questions were again answered in context. Then, the subject worked his way through this second problem on his own. The subject acted as the system operator and entered the commands. After the subject entered his strike timing decisions, he was asked whether he had any additional procedural questions. Then, the formal data collection started.

Data Collection

The subject then proceeded with the eight test problems in the predetermined random order. The sequence was initiated by the administrator who called from the system one of five experimental versions of ASTDA. The version called was appropriate to the evaluative condition. After the subject started, he was left on his own. The administrator stationed himself at the back of the partitioning screen where he could unobtrusively observe the commands entered and displays called. The specialized testing computer routine which was used to present the problems also created a data file for each problem. The file stored timing factors, all the commands entered by the subject, the subject's strike timing rank order decisions, and his confidence and difficulty judgments. The routine measured and stored the time between a command and when the tables and graphs were fully displayed and then measured the time to the next command. In addition, the sequential relationship of the commands was preserved. The routine also asked the subject, at the end of each problem, if he wished to continue with the next problem. A "yes" entry called the next problem. When all eight problems were completed, the routine informed the subject that the formal data collection was finished. The net result was that the subject could work at his own pace without administrator intrusion from data collection start to completion.

After Data Collection Interview

After each subject completed the eight problems, an interview was administered. The interview, generally, attempted to obtain a qualitative evaluation of all aspects of the aid including usefulness and workability. The interview was semistructured in nature and inquired into three specific topic areas. The first attempted to obtain the data required for an assessment of the aid's utility. This assessment was implemented by a multi-attribute utility analysis. The second part of the interview consisted of an evaluation of the usefulness of the aid and its components. The final part was directed toward an evaluation of the organization and content of the display and control systems.

III. RESULTS

The analysis of the emergent data proceeded in an orderly manner. First, a criterion (validity) analysis was completed. This analysis sought to determine the degree of agreement between the two available criteria--expert panel judgments and expected utility as computed by the ASTDA. Then, the differences between the various conditions were examined. Next the data were examined for learning effects. In addition, a multiple correlation and regression analysis was performed on the dependent variables relative to the information made available. Finally, the interview results were examined.

Criterion Analysis

The criterion analysis sought to establish the relationship, if any, between the two sets of criterion data. One set, the EU criteria data, represented the best predictions of the ASTDA while the second represented the pooled judgment of experienced naval operations personnel about the preferred launch time ranking.

Specifically, the expert panel provided, for each problem, the best two and the worst two strike launch times from a choice of six potential launch times. For each problem, the panelist's joint strike launch times were ranked and paired with the ASTDA calculated expected utility. The assignment of utility to the experts' decisions is illustrated in Table 3 for a typical problem. The top left part of Table 3 contains the six possible strike launch times, 1200 to 1700 hours inclusive, while next to each is the mean and the range of the expected utility calculated by the ASTDA for each time.

The bottom half of the table contains two sets of ranked times and their utilities--on the left those of the ASTDA and on the right the utilities for the panel judgments. The highest expected utility in the example is 43.89 which is associated with a 1200 hours launch. The lowest utility occurs for a 1400 hours launch and was assigned the sixth rank by the ASTDA.

Under the heading "Panel" in Table 3, the ranked judgments of the panel appear along with the associated utility. The result was 32 paired values (4 launch times per problem x 8 problems). The raw score intercorrelation between the two arrays was .91 (Figure 3 (A)). Because of differences in the utility distributions within each problem, the data for each problem were converted to normal deviates (z scores) and the correlation coefficient was again calculated. The resultant correlation coefficient was .47 (Figure 3 (B)). These results indicate at least a moderate, positive relationship between the two measures and support contentions favoring the validity of the ASTDA.

Table 3

Method of Comparing Utility of ASTDA and Expert Judgment Launch Times for a Typical Problem

Launch Time	Utility (ASTDA)		
	Mean	Range	z
1200	43.89	(9.00, 69.75)	1.3794
1300	42.05	(6.85, 68.00)	1.1478
1400	25.16	(-25.58, 61.00)	-0.9786
1500	29.68	(-26.21, 64.00)	-0.4096
1600	28.84	(-29.67, 65.00)	-0.5153
1700	27.98	(-32.67, 65.00)	-0.6236

	Aid				Panel		
	Rank	Time	Utility	z	Time	Utility	z
Best Times	1	1200	43.89	1.3794	1200	43.89	1.3794
	2	1300	42.05	1.1478	1300	42.05	1.1478
Worst Times	5	1700	27.98	-0.6236	1600	28.84	-0.6236
	6	1400	25.16	-0.9782	1700	27.98	-0.9782

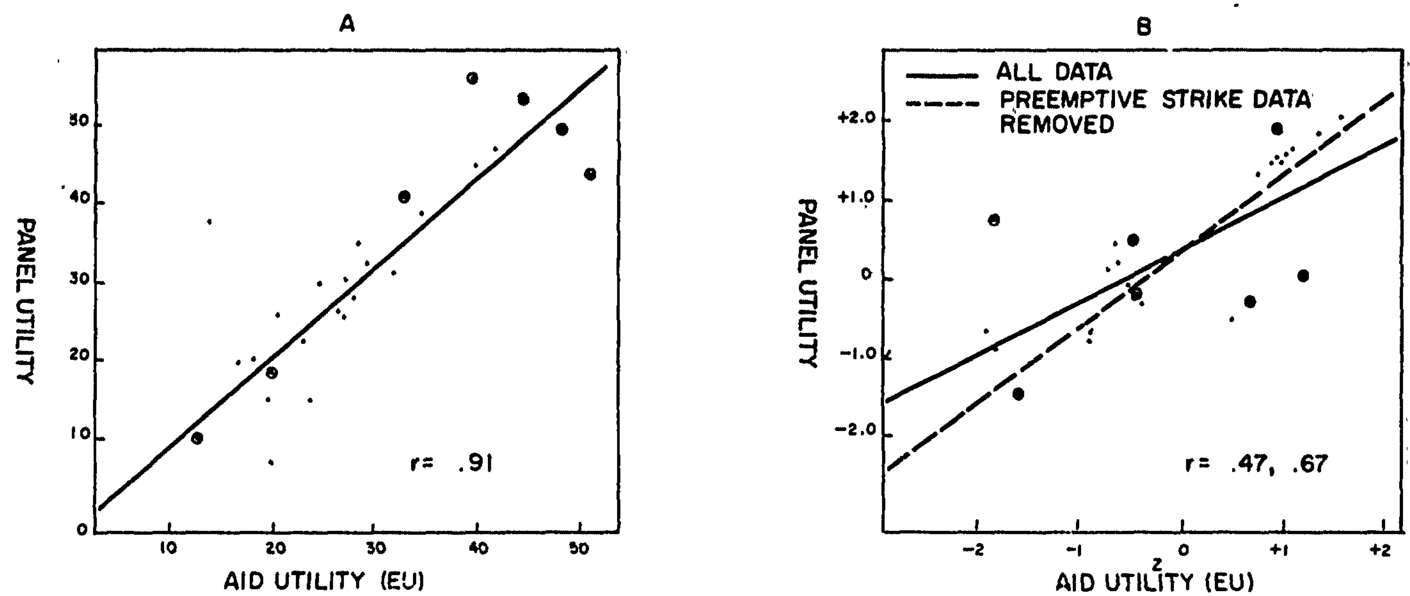

Figure 3. Regression of aid assigned utility (A) and the standardized utility scores (B) on panel utility values.

The panel, for two problems, chose to launch early, preemptive strikes on the basis of military judgment. This judgment encompassed factors not considered by the aid. The decisions to launch preemptive strikes were made regardless of other advantages or disadvantages, e.g., the number of own (Blue) aircraft available, or the number and kind of Orange air defenses. The critical conditions seem to have been a large collection of parked Orange aircraft and intelligence information that an attack was imminent. The panel thought a preemptive, early strike was worth the additional loss of own men and equipment. However, the ASTDA does not include such considerations within its logic. This limitation has implications for aid design and is discussed later. Therefore, it cannot differentially predict the utility of preemptive strikes. Moreover, for the problems involved, the preemptive strike times were fortuitously undesirable as compared with the other times under consideration. For example, for problem 4, the best time chosen by the panel was evaluated by the aid to have a slightly negative value. In the other problem in which preemptive strikes were decided on by the panel, problem 6, the panel's rankings were virtually the opposite of those yielded by the aid.

Accordingly, the data sets for these two problems were eliminated and the correlation coefficient was again computed employing the normalized data. In Figure 3 (B), the circled points represent the eliminated data. The resultant correlation coefficient was .67.

One may also argue that in actuality only one launch time is possible for a given strike. By this argument, only the first choice becomes relevant. Accordingly, only first choices were intercorrelated. The resulting correlation coefficient was .99 ($N = 8$).

If the panel derived data are assumed to represent the criterion to be predicted by the ASTDA, then these correlation coefficients represent validity estimates for the aid itself, i.e., the relationship between the expert's judgments and the aid's prescriptions concerning the problems. The relationship appears positive and moderate in magnitude. This relationship increased dramatically, when only problems which were solved by the panel without preemptive strikes were considered and a further increase was demonstrated when only first choices were considered.

Aid Conditions

Ranked Difference Scores

To analyze the differences among responses as a function of the various aid conditions (see Exhibits I and II), two indexes were used. The first index reflected the difference between the rankings produced by the aid (utility rankings) and those assigned by the subjects for the launch times

associated with each problem. The second index represented a measure of the difference between the panel's rankings of the launch times for each problem and the rankings of the individual subjects.

Exhibit IV illustrates the calculation of ranked difference scores. In calculating these scores, only the ranks 1, 2, 5, and 6 were used because of limits on the data obtained from the expert panel. In the Exhibit IV example, the first ranked time by the aid was 1200 hours while the subject ranked 1200 hours fifth, a difference of four. The four represents the first value of the ranked difference score. The aid ranked 1300 hours as second best while the subject assigned 1300 hours to the first rank; therefore, there was a difference value of one. This became the second value to enter the ranked difference score. For the fifth ranked time, there was a one rank difference between the aid and the subject. In the ee sixth ranked time, there was a two rank difference. The numbers one and two were entered into the ranked difference score for the third and fourth values (fifth and sixth ranks) respectively. Summing across the four difference values gives the ranked difference sum (8) against the aid (expected utility) criterion.

A parallel technique was employed to calculate ranked difference sums using the panel judgments as the criterion. This is illustrated in the bottom half of Exhibit IV where the sum of the difference scores for the subject on the individual problem is six.

The ranked difference scores are an inverse measures of agreement with lower ranked difference scores indicating better agreement. The scores, as calculated, possess a range of zero to 16. A zero rank difference indicates perfect agreement and a rank difference score of 16 indicates a ranking of the launch times by the subject inversely to that of the criterion.

Variance Analysis of Difference Scores Using the Aid as the Criterion

The ranked difference scores based on the ASTDA's utility values as the criterion were subjected to a two (problem difficulty) by five (aid levels) by two (subject experience) analysis of variance. The results of this analysis are presented in Table 4.

The results of the variance analysis indicated statistically significant variance due to the aid level and the problem difficulty main effects. The aid level by problem difficulty interaction was also statistically significant.

The mean values for the various conditions are summarized below:

Aid Level	Mean
Full Aid (A1)	2.20
Utility (A2)	2.95
Outcome (A3)	2.96
No Uncertainty (A4)	2.74
No Aid (A5)	6.60

Exhibit IV

Example of Calculation of Ranked Difference Score

	Rankings						
	1	2	3	4	5	6	Ranked Difference Score
Subject	1300	1600	1500	1400	1200	1700	
Aid (EU)	1200	1300			1700	1400	
Difference	4	1			1	2	Σ = 8
Experts	1300	1200			1600	1700	
Difference	0	3			3	0	Σ = 6

Table 4

Variance Summary for the Ranked Difference Scores--EU Criterion

Source	Sum of Squares	df	Mean Square	F
Between	2437.50			
Experience (E)	0.60	1	0.60	0.03
Aid Levels (A)	1198.12	4	299.53	15.76*
E x A	49.49	4	12.37	0.65
Error: Subjects within groups	950.40	50	19.01	
Difficulty (D)	85.85	1	85.85	6.62*
D x E	5.85	1	5.85	0.45
D x A	133.49	4	33.37	2.57*
D x E x A	13.70	4	3.43	0.26
Error: D x Subjects within groups	648.23	50	12.96	

*p = 0.05

Table 5

Variance Summary for the Ranked Difference Scores--Panel Criterion

Source	Sum of Squares	df	Mean Square	F
Between	1907.15			
Experience (E)	21.67	1	21.67	1.11
Aid Levels (A)	159.83	4	39.96	2.05**
E x A	14.28	4	3.57	0.18
Error: Subjects within groups	974.37	50	19.49	
Difficulty (D)	0.07	1	0.07	0.01
D x E	0.13	1	0.13	0.01
D x A	192.88	4	48.2[illegible]	4.67*
D x E x A	27.78	4	3.95	0.67
Error: D x Subjects within groups	516.13	50	10.32	

*p = 0.05

**approaches p = 0.05

Difficulty	Mean
Hard	3.83
Easy	3.07
Experience	
Operational Experience	3.46
No Operational Experience	3.53

The lowest mean score, 2.20, was observed in the full aid condition while slightly higher (poorer) ranked difference scores (2.95, 2.96, and 2.74) resulted from the utility, the outcome, and the no uncertainty conditions, respectively. The highest, most deviant difference score, 6.60, was observed in the no aid control condition. A Newman-Keuls a posteriori comparison of the means using the Studentized range statistic indicated no statistically significant differences among the first four conditions but that the no aid condition differed significantly from the others.

In addition to aid levels, the difficulty of the scenario problems also produced systematic variance differences in the ranked difference scores. Difference scores tended to be higher (poorer) when working with the hard scenario problems. The effect of difficulty also significantly interacted with aid levels. In Figure 4, within aid levels, the mean ranked difference scores for the hard scenario problems are consistently higher than those for the easy problems across the first four aid levels. In the fifth aid level, the no aid level, this relationship is reversed. A comparison of the means through the Newman-Keuls test indicated no statistically significant differences across aid levels for the hard problems. However, for the easy problems, the mean ranked difference scores from condition A5, the no aid condition, differed significantly from the means of all the other conditions except the utility condition.

A comparison by aid levels across difficulty levels did not indicate any systematic differences.

It appears that, in terms of the ranked difference scores here involved, there was little consistent tendency for the subjects to rank order the launch times different in any condition in which all, or parts, of the output displays (BFL, ORAL, ORGL, and EU) were available. That is, not having the outcome displays, or the utility display, or the uncertainty bands did not significantly affect the variance across ranks. Only when the subjects were not given any output information in condition A5 was there any appreciable effect on their choices. It is possible that the information contained in the various displays is positively correlated. Accordingly, increased information volume may have contributed little.

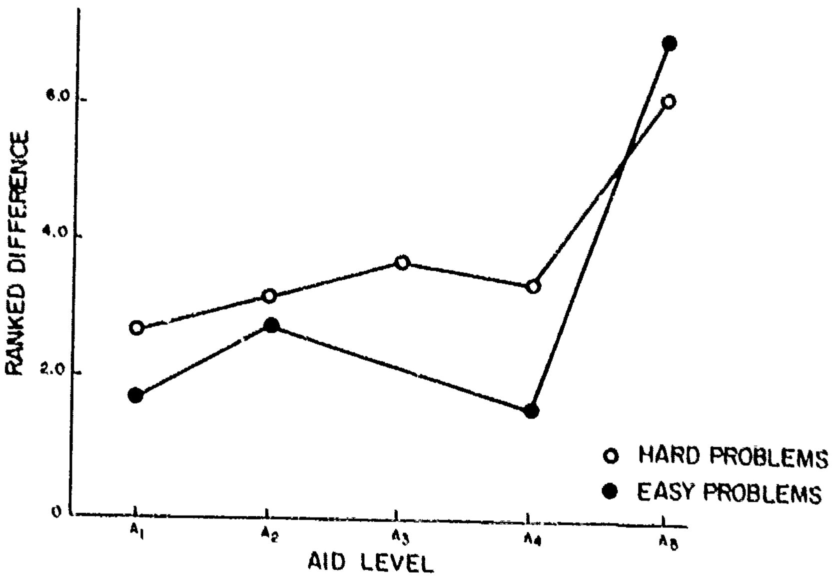

Figure 4. The mean ranked difference scores from the EU criterion for each aid and difficulty level. Aid Levels (A1, A2, A3, A4, and A5) correspond to the full aid, utility, outcomes, no uncertainty, and no aid conditions, respectively.

The distribution of times was also affected by problem difficulty; hard problems led to a wider divergence from the criterion while easy problems produced rankings closer to those of the criterion. However, the difficulty by aid level interaction data suggest that most of the differences associated with difficulty were produced by the no aid condition.

Variance Analysis of Difference Scores Using the Panel's Judgments as the Criterion

The ranked difference scores were also analyzed with the expert's judgments forming the criterion. The only statistically significant variance was attributable to the aid level by difficulty interaction. Generally, the difference scores produced by this analysis tended to be higher than those calculated on the basis of the aid produced expected utility criterion. The mean ranked difference score across all conditions was 3.45 for the expected utility but it was 6.08 when calculated against the expert's judgments. Part of this difference might be due to the inclusion in the analyses of those problems for which the experts chose to launch preemptive strikes. In addition, both of these preemptive problems were classified as easy. This would tend to inflate the ranked difference scores for the easy problems. A summary of a variance analysis of these data is presented as Table 5. Within the analysis of variance, there were few systematic differences across or within conditions. There was a tendency toward significant effects across the aid levels. The variance due to experience or problem difficulty was not statistically significant.

The mean ranked difference scores for each condition were:

Aid Level	Mean
Full Aid (A1)	5.40
Utility (A2)	5.40
Outcome (A3)	6.31
No Uncertainty (A4)	6.72
No Aid (A5)	6.58
Difficulty	
Hard	6.09
Easy	6.07
Experience	
Operational Experience	6.47
No Operational Experience	6.29

A comparison of the aided conditions with the no aid condition seems to suggest that in the full aid condition the subjects tended to have lower (better) scores than in the no aid condition. The utility condition (A2) yielded scores more like the full aid (A1) and the no uncertainty condition (A4) produced scores more like the no aid condition.

The subjects tended to produce better (lower) rank difference scores in the full aid and the utility conditions than for the other aid levels involved.

The statistically significant interaction effect is plotted in Figure 5.

Regret

The prior analyses depended on the sensitivity of the ranked difference scores to the various evaluation conditions. It is possible that scoring on the basis of launch time rankings may have obscured real differences between strike times and not accurately weighed the consequences of differences among the rankings. Consider the case in which the best strike time has a utility of 50 and the second best time has a utility of 49. Ranked that way by the subject and the criteria, the result is a zero ranked difference score. If the subject reversed the times so that he assigned a rank of one to the time with a utility of 49 and a two to the time with a utility of 50, then a one would be generated as the ranked difference score. But is a difference of one in utility units equivalent to a rank difference of one? If the second best time according to the aid had a utility of 40, would such a difference represent an important difference? If a subject ranked as best the time with the 40 utility and as second best the time with the 50, should this inversion be given the same weight as the difference between 49 and 50?

It seemed possible that a "regret analysis" would allow for a more sensitive evaluation of the differences between utilities. The regret score is defined as the difference between the utility associated with the time specified by the aid and the utility associated with the time chosen by the subjects. If the best time predicted by the aid had a utility of 50 and the best time chosen by a subject had a utility of 45, then the difference between these utilities, 5, represents the regret score. Regret scores were calculated in order to assess differences in utility value between times indicated by the aid and those chosen by the subjects. The regret scores were only calculated for the best launch time because it was thought that second or third best times represented academic issues of minimum consequence to the operational situation.

Variance Analysis of Regret Scores

The regret scores for the expected utility criterion were subjected to a two by five by two variance analysis. The differences between predictions made by the aid and the subjects choice times were first examined. The analysis indicated significant systematic variance across aid levels, difficulty levels, and their interactions. These differences parallel those which resulted from the analysis of the ranked difference scores. In addition, there was also a tendency toward significant differences across experience levels (operationally experienced/operationally inexperienced). The summary of this variance analysis is presented as Table 6.

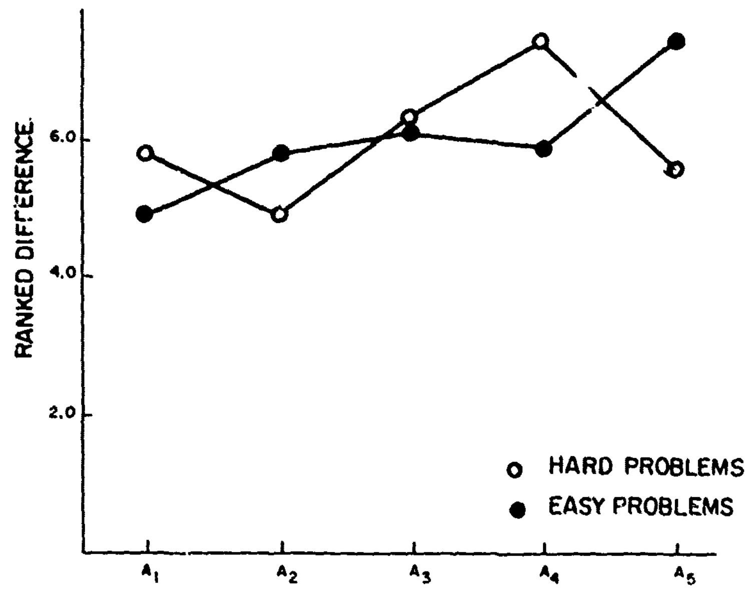

Figure 5. Mean ranked difference scores from the panel criterion for each aid and difficulty level. Aid levels (A1, A2, A3, A4, and A5) correspond to the full aid, utility, outcome, no uncertainty, and no aid conditions, respectively.

Table 6

Variance Summary for the Regret Scores--EU Criterion

Source	Sum of Squares	df	Mean Square	F
Between	4718.67			
Experience (E)	40.84	1	40.84	2.67*
Aid Levels (A)	1620.11	4	405.03	26.48**
E x A	70.69	4	17.67	1.16
Error: Subjects within groups	764.92	50	15.30	
Difficulty (D)	360.45	1	360.45	20.98**
D x E	5.00	1	5.0	0.29
D x A	962.46	4	240.61	14.01**
D x E x A	35.35	4	8.84	0.51
Error: D x Subjects within groups	858.67	50	17.17	

*tendency toward significance, p = 0.109

**p = 0.05

The mean regret scores for the various conditions were:

Aid Level	Mean
Full Aid (A1)	0.51
Utility (A2)	1.58
Outcome (A3)	1.09
No Uncertainty (A4)	0.97
No Aid (A5)	5.57
Difficulty	
Hard	1.07
Easy	2.81
Experience	
Operational Experience	1.64
No Operational Experience	2.23

The scores from the first four aid levels tended to be rather similar with the lowest (best) score, 0.51, resulting from the full aid condition and the highest (worst) score, 1.58, resulting from the utility condition. The mean regret score for the no aid condition (5.57) was considerably higher (worse) than that for the remaining levels. A Newman-Keuls test indicated that the mean regret score for the no aid condition varied significantly from the rest while no statistically significant differences occurred among the means of the other conditions.

The operationally experienced group's mean regret score was 1.64. This value was substantially lower (better) than the inexperienced group's mean regret score of 2.23. The mean difference represents a tendency towards statistical significance.

The difficulty of the scenario problems also affected the regret scores. When the problems were hard, the regret scores were fairly low with a mean of 1.07 and when they were easy the scores were significantly higher with a mean of 2.81.

In addition to the significant main effects, the first order interaction was statistically significant. The interaction effect is shown in Figure 6 which also indicates that some form of aiding decreased regret scores considerably--especially for easy problems.

The easy and the hard problems produced significant differential effects within the aid levels. An examination of the means indicated that in the full aid, the outcome, and the no uncertainty conditions, the regret scores varied little across difficulty levels. However, in the other two conditions, utility and no aid, there were noticeable increases in the regret scores of the easy problems as compared with the hard problems. A

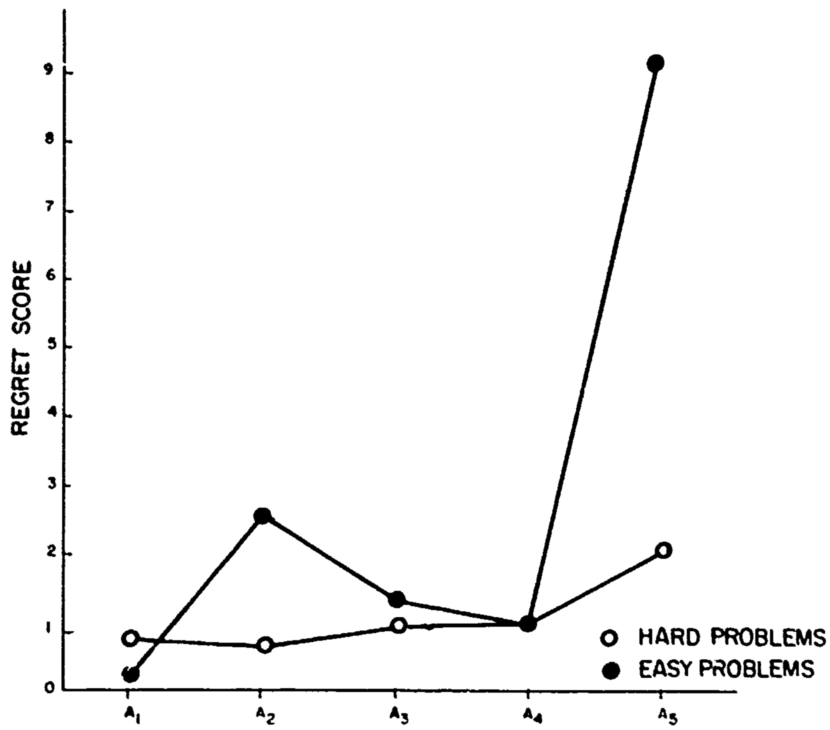

Figure 6. Mean regret score for the EU criterion for each aid and difficulty level. Aid levels (A1, A2, A3, A4, and A5) correspond to the full aid, utility, outcomes, no uncertainty, and no aid conditions, respectively.

Newman-Keuls analysis indicated no systematic mean differences among hard problems across aid levels. On the other hand, for the easy problems the mean regret scores from the no aid condition (A5) was significantly higher than the others. A comparison across difficulty within each aid level indicated that the only statistically significant difference was in the no aid condition.

These effects of difficulty seem to be reversed from the trend for the ranked difference scores. Difficulty decreased the regret scores for the hard problems and increased it for the easy problems, while the opposite effect was noted for the ranked difference scores. This seeming contradiction may be resolved by comparing the aid level by difficulty interaction data. The aposteriori comparison of means indicated that the only significant effect of difficulty for both the regret scores and the ranked difference scores was due to the no aid, easy problem condition. Accordingly, we believe the seeming reversal to be due to the no aid condition. As such, the difference is artifactual or not of immediate interest to an evaluation of the ASTDA.

Learning

It is possible that as a subject worked his way through the problems, he may have learned some important aspect about the use of the aid, the variables involved, and the context. Such learning might be expressed as an increasing approximation to the predictions of the aid.

To evaluate this possibility, the successive ordering of problems for each subject was recovered and the problem set was divided into halves. The early half consisted of the first four problems the subject completed and the late half consisted of the last four. The performance measures used to evaluate learning effects were the ranked differences scores from the aid and from the experts. A separate analysis of variance was completed for each set of criterion data.

Results

A summary of the variance analysis for learning effects employing the aid computed utility values as the criterion is presented as Table 7 and a parallel summary employing the panel's judgments as the criterion is presented as Table 8. The analysis employing the expert judgment criterion failed to indicate any statistically significant differences. The analysis employing the aid calculated utility criterion indicated a statistically significant main effect due to aid levels and a tendency towards a significant three way interaction. The main effect result is not pertinent to the learning question. The interaction data are presented in Figure 7. Scores tended to decrease (improve) in two conditions for the experienced subjects and in three conditions for the inexperienced. Also, the magnitude of the variance within aid levels was much more pronounced for the inexperienced subjects.

Table 7

Variance Summary for the Learning Data--EU Criterion

Source	Sum of Squares	df	Mean Square	F
Between	3010.94			
Experience (E)	2.00	1	2.00	0.10
Aid Levels (A)	1054.04	4	263.51	13.37**
E x A	59.42	4	14.86	0.75
Error: Subjects within groups	985.10	50	19.70	
Learning (L)	5.42	1	5.42	0.40
L x E	8.27	1	8.27	0.61
L x A	99.01	4	24.75	1.83
L x E x A	122.1584	4	30.54	2.26*
Error: L x Subjects within groups	967.52	50	13.51	

*p = 0.1

**p = 0.05 (difference not relevant to question of learning)

Table 8

Variance Summary for the Learning Data--Panel Criterion

Source	Sum of Squares	df	Mean Square	F
Between	2544.46			
Experience (E)	1.32	1	1.32	0.08
Aid Levels (A)	83.32	4	20.83	1.24
E x A	59.43	4	14.86	0.88
Error: Subjects within groups	839.50	50	16.79	
Learning (L)	5.37	1	5.37	0.19
L x E	7.82	1	7.82	0.28
L x A	53.37	4	13.34	0.48
L x E x A	107.29	4	26.82	0.97
Error: L x Subjects within groups	1387.03	50	27.74	

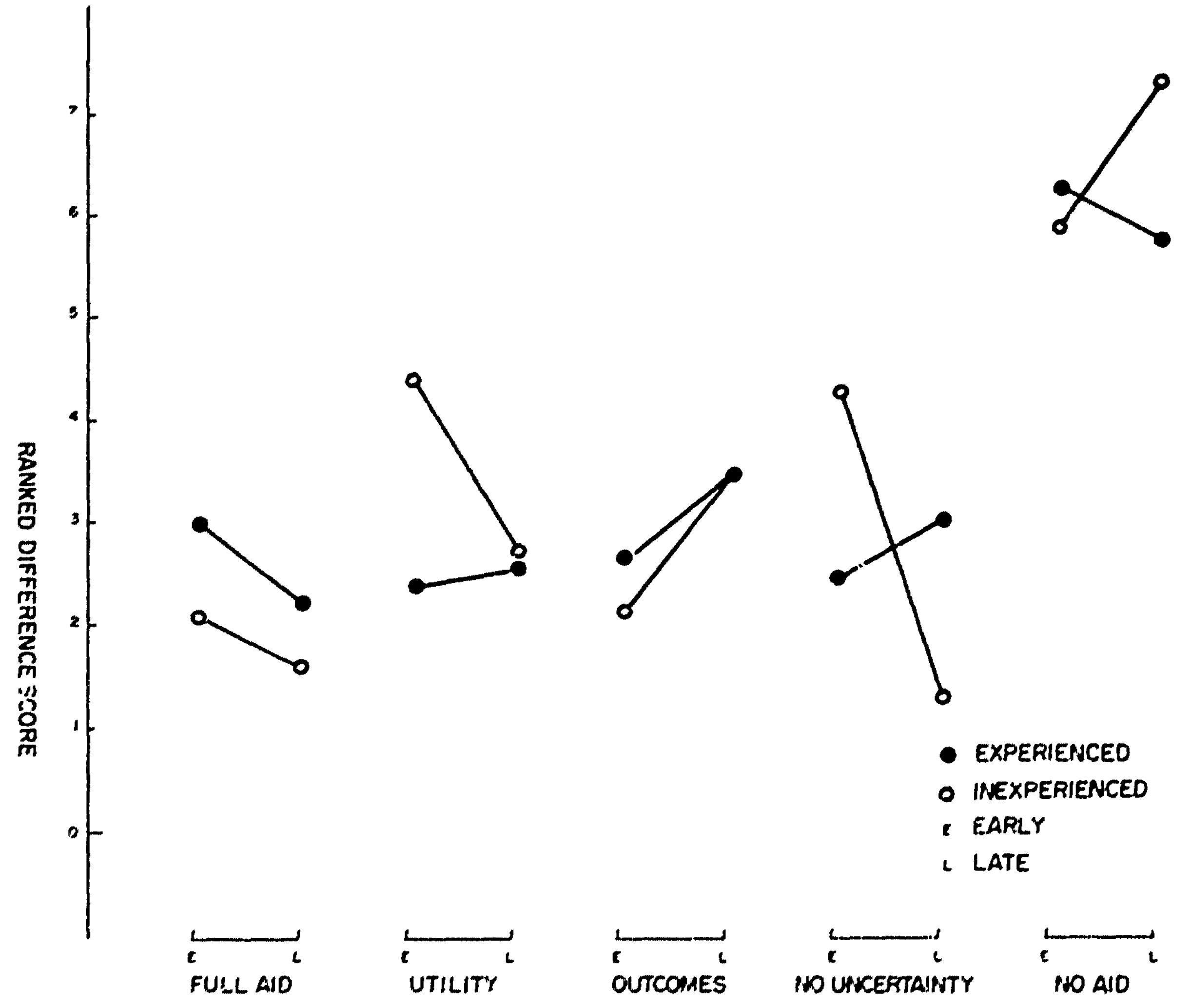

Figure 7. Interactions among early (E), late (L), aid levels, and experience levels.

Given the design of the aid and the kinds of information it supplies, there was little to learn about the use of the aid or interpretation of its output. There may have been an increase in confidence in the aid's assessments with time. However, this is not the factor of interest here. The results suggest that there was little difference in the scores across successive (early versus late) problems.

Policy Capturing

For each of eight problems, each subject was required to decide on the ranking of six possible launch times. To derive his decision, the subject had two sources of information available. "Objective" information was supplied by the aid. (Although objectivity does not necessarily imply factuality or accuracy.) The subject also relied on his personal intuition which might be called his cognitive-operational-emotive perception, strategies, or schemes.

We are concerned here only with the objective information because of our emphasis on evaluating the aid itself. An understanding of what aid produced information the subjects emphasized and what weight they attached to it would provide insight into their decision process. Such information might also provide insight into design requirements for such aids.

The relationship between the subjects' choice of strike launch times and the objective information available was analyzed. This analysis was performed by use of the multiple regression technique. Such an approach has been termed "policy capturing" by others because it essentially reveals the policy followed by the subjects in deriving their decisions (Christal, 1968).

One dependent and eleven independent variables were included in the analysis. The values of the dependent variable were the rankings of the best two and the poorest two launch times. The values assigned to the independent variables each represented a value derived from the various displays: the BFA, ORAD, WAT, WAC, BFL, ORAL, and EU displays. The information from the other two displays, the ORGF and ORGL was separated into four sets of data: the Orange ground defenses (ORGD), the Orange ground targets (ORGT), the Orange ground defense losses (ORGL), and the Orange targets destroyed (ORTD). The assignment of values to the dependent and independent variables is illustrated in Table 9 for a hypothetical subject. The two best and the two poorest times are listed on the left of the table. Instead of using the times as the values of the dependent variable, rankings were consistently assigned. A "4" was assigned the first choice and a "1" was assigned to the poorest time.

Table 9

Example of Data Organization for Regression Analyses

Dependent Variable		Independent Variables										
Times	Rank	BFA	ORADR	ORGDR	ORGT	WAT	WAC	BFLR	ORAL	ORGL	ORTL	EU
1700	4	30.9	.032	.053	10.9	.96	.97	.179	13.3	.3	4.7	38.35
1400	3	25.4	.036	.072	10.9	.71	.89	.135	9.0	.2	3.1	10.30
1500	2	28.1	.034	.063	10.9	.80	.91	.147	10.6	.2	3.7	20.90
1200	1	22.3	.042	.094	10.9	.87	.92	.227	9.2	.1	3.3	24.05

BFA: Blue Force Availability
ORADR: Reciprocal of Orange Air Defenses
ORGDR: Reciprocal of Orange Ground Defenses
ORGT: Orange Ground Target
WAT: Weather at the Target
WAC: Weather at the Carrier
BFLR: Reciprocal of Blue Force Losses
ORAL: Orange Air Losses
ORGL: Orange Ground Losses
ORTL: Orange Target Losses
EU: Expected Utility

Corresponding to a particular time chosen by the hypothetical subject, the values of the eleven independent variables are given. These represent the data available to the subject vis-a-vis each time. When the displays did not supply a single reference number for a particular time, as was the case for the ORGT, WAT, WAC, and EU displays, the sum of the different unit types was used as the appropriate value of the dependent variable, e.g., the BFA value of 30.9 for 1700, in Table 9, is the sum of the number of BF1s, BB1s, and BB2s (various aircraft types) available at that particular time. A positive relationship existed for most independent variables between the perceived goodness and the value of the independent variables. For example, as the number of blue aircraft available (BFA) increased, the perceived goodness to blue similarly increased. However, for the ORAD, ORGD, and BFL displays, the reverse was true. Accordingly, the reciprocals of the value for each of these independent variables was entered into the analysis.

A separate multiple regression analysis was conducted for various conditions, i.e., by experience of subject, problem difficulty, aid levels, and combinations. An overall analysis, with data collapsed across all conditions, was also completed. Each analysis was step wise. Cut off criteria of $F = 2.00$ and a tolerance of 0.1 were established.

Overall Analysis

In the overall regression analysis the data were collapsed across aid, difficulty, and background levels. The results are presented in Table 10. The zero order correlation matrix for each multiple regression analysis is found in the Appendix to this report).

The first four variables entered into the equation represented ASTDA input display information and the next three represented output display information. The first variable to enter the equation, the WAT, yielded an R of .52 and accounted for 27 percent of the total variance or about 70 percent of the predictable variance. Each of the other variables, as they were entered, accounted for a lesser amount of variance. The range for the successive variables was from about 4 percent of total variance for the ORAD reciprocal to less than .5 percent for the BFA. Overall, the multiple correlation between the decisions of the subjects and the variables entered into the equation was .61, accounting for 38 percent of the total variance. One may possibly assume that the remaining variance can be accounted for in some part by the cognitive-operational-emotive experience of the subjects. Generally, the weights and order of entry of variables into the equation seem to be reasonable. The strongest influence on choice was apparently exerted by the WAT and the ORAD. Other variables tended to influence the dependent variable very little. Note also that the first four variables to enter were based on enemy posture/conditions rather than on the friendly (blue) situation.

Experience

A separate regression analysis was completed in which the experience of the subjects was fractionated. The resultant data are presented in Table 11. Background did not seem to produce any important observable differences either between groups or as compared to the data from the overall analysis. The Rs, R^2s, R^2 changes, B weights, and constants were remarkably similar. The variables and their order of entry from the separate experienced and inexperienced groups were similar to those from the overall analysis. A comparison of the results for the experienced and the inexperienced groups indicates only minor differences in the order in which the variables entered the equations. The variance analysis also failed to indicate consistent differences across experience levels. It seems that, at least for the conditions of the present evaluation, operational experience was not a substantial influence on either accuracy or on the decision making policy.

Difficulty

The data were separated by problem difficulty assignment and similarly analyzed. The analysis of the hard problems (Table 12) indicated that the first two variables to enter the equation were the BFL reciprocal and the EU. This was the first regression analysis in which the variables derived from ASTDA output information entered early and the first time that EU entered the equation at all. In the analysis of the data from the easy problems, the ordering of the variables was similar (but not congruent) with that observed in the overall analysis. That is, the first two variables entered (the WAT and the ORGD reciprical) were from ASTDA input displays.

There were also substantial differences in the amount of variance accounted for between the easy and the hard problems. In all, five variables were entered into the multiple regression equation for the hard problems and a multiple R of .62 accounting for 30 percent of the variance was produced. However, the multiple R for the easy problems was higher, (.75) and accounted for 57 percent of the total variance.

This seems to suggest that differences existed in the human information processing for the easy and the hard problems and in the consistency with which the information was used. On the one hand, for the hard problems, the data suggest that the subjects tended to make choices in line with specific sources of output information--the BFL reciprocal and the EU--which together accounted for 28 percent of the variance. Then, they apparently qualified their choices by considering specific input information supplied by the ORGD, ORGT, WAT, and WAC displays. On the other hand, when working with easy problems, the major correlates of the decisions seem to have been input information from the WAT and the ORGD, which together accounted for 52 percent of the total variance. The decisions appear to be further qualified by considering other sources of both input and output information: the ORGT, WAC, BFL, ORGL, and ORAL.

Table 10

Results of Overall Regression Analysis

Order of Variable Entering Equation	Multiple R	R^2	R^2 Change	B Weight	A Constant
WAT	.5185	.2688	.2688	2.00	
ORADR*	.5548	.3078	.0389	7.25	
ORGT	.5639	.3180	.0101	-0.03	
ORGDR*	.5722	.3274	.0094	-33.97	
BFLR*	.5908	.3490	.0217	17.73	
ORTL	.6070	.3684	.0194	0.18	
ORGL	.6101	.3723	.0038	-1.50	
BFA	.6131	.3759	.0037	-0.04	
					-1.10

*Variables entered as reciprocals; R = reciprocal

Table 11

Results of Regression Analysis by Background

Experienced

Order of Variable Entering Equation	Multiple R	R^2	R^2 Change	B Weight	A Constant
WAT	.5138	.2640	.2640	1.70	
ORADR*	.5508	.3034	.0394	6.14	
ORGT	.5636	.3177	.0143	-0.03	
ORGDR*	.5705	.3255	.0077	-34.30	
BFLR*	.5910	.3493	.0238	18.00	
ORTL	.6073	.3688	.0195	0.21	
ORGL	.6112	.3736	.0048	-1.61	
BFA	.6133	.3751	.0025	0.03	
					-.62

Inexperienced

Order of Variable Entering Equation	Multiple R	R^2	R^2 Change	B Weight	A Constant
WAT	.5233	.2738	.2738	2.30	
ORADR*	.5589	.3147	.0336	8.40	
ORGDR*	.5670	.3215	.0090	-33.70	
BFLR*	.5862	.3437	.0022	17.49	
ORGL	.6029	.3636	.0199	0.15	
ORTL	.6079	.3695	.0059	-1.41	
ORGT	.6100	.3721	.0026	-0.02	
BFA	.6142	.3773	.0051	0.05	
					-1.46

*Variables entered as reciprocals; R = reciprocal

Table 12

Results of Regression Analysis by Problem Difficulty

Hard Problems

Order of Variable Entering Equation	Multiple R	R^2	R^2 Change	B Weight	A Constant
BFLR*	.4426	.1959	.1959	28.75	
EU	.5303	.2812	.0853	0.10	
ORGDR*	.5537	.3066	.0253	-74.53	
ORGT	.6060	.3608	.0542	-0.12	
WAT	.6162	.3798	.0190	-5.63	
WAC	.6174	.3812	.0015	1.56	
					5.82

Easy Problems

Order of Variable Entering Equation	Multiple R	R^2	R^2 Change	B Weight	A Constant
WAT	.6724	.4521	.4521	5.63	
ORGDR*	.7229	.5226	.0706	-25.70	
BFLR*	.7410	.5492	.0265	6.77	
ORGT	.7450	.5551	.0059	0.04	
ORGL	.7494	.5616	.0065	-2.34	
WAC	.7510	.5641	.0025	1.35	
ORAL	.7521	.5657	.0016	0.07	
					-3.43

*Variables entered as reciprocals; R = reciprocal

Aid Levels

Table 13 summarizes the regression analyses completed in regard to aid levels. There was a sharp dichotomy between aided conditions (in which all or part of the output displays were available) and the no aid condition. In the no aid condition, only minimal variance ($R = .37$; variance accounted for = 13 percent) was accounted for by the multiple regression equation. This seems to suggest little consistency among these subjects in how they solved the problems producing a distribution of choices that was almost uncorrelated with any of the independent variables included in this analysis.

This was not the case for the aided conditions where the variance accounted for was much higher. The variance accounted for was highest in the full aid conditions in which 54 percent of the total variance was identified. Lower amounts of variance were accounted for in the utility, outcome, and no uncertainty conditions, (46, 44, and 46 percent respectively). Consistently, the first two variables entering the equation were the WAT and the ORAD. These two variables, together, accounted for between 76 and 82 percent of the total variance that was accounted for.

Other regression analyses were completed on the data. These analyses involved aid by background, aid by difficulty, background by difficulty, and aid by background by difficulty. The results produced multiple regression equations which were very similar to those reported.

Discussion of Regression Analyses

By and large, the most powerful single correlate of choice was the weather at the target followed by information about the enemy air defenses. This generalization is mitigated when the effects of difficulty are considered. For hard problems own losses and expected utility were strongly related to choice. Hence, difficulty level seems to act as a moderating variable on decision policy.

It seems that two fairly distinct sets of information were used when solving easy as compared with hard problems. Emphasis in solving easy problems was based primarily on input information. For hard problems, output information played a somewhat greater role. The result was somewhat variant regression solutions for the two problem types.

Hence, the aid configuration which is best for one problem difficulty level may not be best for another difficulty level. The analyses of the aid levels indicated differences between aided conditions and no aid conditions. There were differences among the variables entering the equation and in the variance accounted for. The two variables which were most highly correlated with the decisions, WAT and ORAD, were available in all

Table 13

Results of Regression Analyses for Various Aid Conditions

Full-Aid

Order of Variable Entering Equation	Multiple R	R^2	R^2 Change	B Weight	A Constant
WAT	.6094	.3714	.3714	1.34	
ORADR*	.6513	.4241	.0527	7.82	
ORGT	.6677	.4458	.0217	-0.04	
ORGDR*	.6752	.4559	.0101	-43.05	
BFLR*	.7058	.4982	.0423	23.84	
BFA	.7278	.5296	.0314	0.06	
ORGL	.7316	.5351	.0056	-1.80	
ORTL	.7343	.5391	.0039	0.18	
					-1.19

Utilities

Order of Variable Entering Equation	Multiple R	R^2	R^2 Change	B Weight	A Constant
WAT	.5525	.3052	.3052	1.16	
ORADR*	.6045	.3654	.0602	-0.39	
ORGL	.6244	.3899	.0245	-4.29	
ORGT	.6331	.4043	.0110	-0.03	
WAC	.6372	.4060	.0051	-1.08	
ORGDR*	.6403	.4100	.0040	-34.64	
EU	.6565	.4310	.0209	0.05	
BFLR*	.6809	.4636	.0326	13.64	
					2.81

Outcome

Order of Variable Entering Equation	Multiple R	R^2	R^2 Change	B Weight	A Constant
WAT	.5755	.3313	.3313	0.98	
ORADR*	.6016	.3620	.0307	-10.28	
ORGDR*	.6142	.3773	.0153	-37.83	
BFLR*	.6321	.3996	.0223	18.32	
ORTL	.6645	.4416	.0421	0.36	
					0.19

No Uncertainty

Order of Variable Entering Equation	Multiple R	R^2	R^2 Change	B Weight	A Constant
WAT	.5677	.3223	.3223	1.55	
ORADR*	.5874	.3451	.0228	-7.	
ORGDR*	.5999	.3599	.0148	-43.	
BFLR*	.6243	.3898	.0298	19.86	
ORTL	.6687	.4471	.0573	0.42	
ORGT	.6745	.4547	.0076	-0.03	
WAC	.6778	.4593	.0045	-1.36	
					1.29

Table 13 (cont.)

No Aid

Order of Variable Entering Equation	Multiple R	R^2	R^2 Change	B Weight	A Constant
WAC	.2891	.0835	.0835	2.14	
BFLR*	.3294	.1085	.0249	8.18	
ORGDR*	.3584	.1284	.0199	-11.46	
ORGL	.3658	.1338	.0054	-1.11	
					0.29

*Variables entered as reciprocals; R = reciprocal

conditions. That WAT and ORAD were not used consistently in the no aid condition might be understood on the basis that the subjects in the no aid condition did not have the parts of the aid that served feedback functions. Parhaps, WAT and ORAD exerted their influence in the aided conditions because when output information was available the effects of these variables could be clearly understood. That is, the output may have served a feedback function, sensitizing subjects to the effects of the input information. Sensitive to these effects, the subjects may have placed their emphasis on WAT and ORAD. The lack of the feedback mechanism in the no aid condition may have served to prohibit the subjects from being sensitive to the variables. The subjects may have assessed the situation by some subjective criterion resulting in distributions of choices which were not strongly correlated with any of the independent variables included in the regression analyses.

The concept of the aid as a feedback mechanism may also account for the greater variance accounted for in the full aid condition as compared with the partially aided conditions. The three types of information available (the outcomes, expected utilities, and statements of variability in the predictions) may have complemented one another producing more sensitive feedback functions and hypotheses than occurred with less complete combinations of information. Again, this may have enhanced the tendency to correlate choices with certain specific classes of information, either WAT and ORAD, or BFL and EU, depending on the nature of the problem.

Examination of Merit

The merit of the ASTDA may be specified as an estimate of decision quality when the full aid was used as compared with the decision quality in the no aid condition. Decision quality may be defined as the relationship between decisions made by the evaluation subjects and those made by the expert panel in each condition. Specifically, the correlation (and the variance accounted for) between the experts' judgments and the subjects' choices in the fully aided condition versus the no aid condition may be employed to yield a measure of merit for the ASTDA.

To this end, a number of product moment correlation coefficients were calculated. These were based on the utility value associated with the best launch time for each problem specified by the experts and that chosen by the subjects. The resultant correlation coefficients are shown below:

All Problems

	Full Aid	No Aid
r	.39	.27

Without Preemptive Problems

	Full Aid	No Aid
r	.87	.33
Mean	.71	.30

All data were normalized prior to calculating the correlation coefficients and the mean r values were calculated with the normal z transformation. The mean correlation between the fully aided condition and the experts was .71--accounting for 50 percent of the variance. The mean correlation between the no aid and the experts was .30--accounting for nine percent of the variance. This suggests a 5 to 1 ratio which reflects the difference between the variance accounted for when using the aid and when not using the aid in relation to the expert opinion criterion. Stated alternatively, use of the ASTDA increased decision validity by a factor of five.

Difficulty and Confidence Rating Data

After a subject recorded his strike launch time choices, he was asked about his confidence in his decisions and how difficult it was to arrive at the decisions. As indicated in the earlier section on Problem Selection, difficulty was defined in terms of the spread of possible outcomes across potential strike times. A rather strong negative correlation between the confidence and the perceived difficulty ratings ($r = -.62$) was evidenced. However, no correlation between the a priori difficulty values and either the confidence ratings ($r = -.15$) or the perceived difficulty ratings ($r = -.03$) was evidenced. This lack of any relationship with the previously defined difficulty was surprising because the variance and the regression analyses showed clearly differential effects of difficulty. It seems that the subjects did not perceive the problems in which the BFL and EU were the major correlate of choice to be more difficult. Possibly, their perception of the situation was one of more confusion or one which demanded more consideration but not difficulty per se. However, the subjective report of difficulty does not seem to be associated with our "objective" measure.

Multiattribute Utility

Method

An attempt was made to evaluate further the perceived utility of the aid by assessing how closely the aid achieved its goals. Six ASTDA goals were developed. They are listed in Table 14. One requirement of the multiattribute utility analytic technique (Edwards, 1971) is the relative

Table 14

ASTDA Goals and Weights

	Goal	Weight
1)	To provide a system to assist in the derivation of the best possible time to launch an air strike.	35
2)	To provide a method for structuring and organizing available information pertinent to the strike timing decision.	5
3)	To provide, given available data, the possible results of various strike launch time decisions.	15
4)	To provide information about the trade-offs (e.g., own or enemy losses) relative to various strike time decisions.	25
5)	To provide a criterion against which strike timing decisions can be evaluated or verified.	13
6)	To support the decision maker so that various strike timing decisions can be made more quickly/accurately.	7

importance of the goals (elements) being considered. Judgments of goal importance and the assignment of goal weights were completed by two of the Applied Psychological Services' staff members who were involved in the ASTDA evaluation. Each independently distributed 100 points among the goals to reflect his judgment of the importance of each goal. The weights were then compared, discussed, adjusted, and mutually agreed on. The weights are included in Table 14.

Each subject who participated in the study was asked in an after evaluation interview to assign a rating along a "0" to "100" scale on the extent to which the aid achieved each goal. By multiplying the weight of a goal and the mean of the rating on the extent to which the aid achieved the goal, a utility measure for the aid in reference to that goal was obtained. This procedure was completed separately for each goal. The resultant utility values are presented in Table 15. The top portion of Table 15 presents the data collapsed across conditions, and the lower portion presents the results by experience, aid level, and background by aid level.

The maximum value that could be attained relative to each goal and the marginal total are shown in parenthesis at the top of Table 15. As can be seen, goals 2 and 6 ere closely achieved by the aid. They were judged to have been about 91 percent and 89 percent satisfied, respectively. The other goals (1, 3, 4, and 5) were rated as 85, 81, 78, and 80 percent satisfied, respectively. These values seem rather impressive.

Comparison across experience levels indicates only minor total differences due to this effect. There were only minor differences in ratings within goals of about 3 to 7 percent. Exceptions were the 9 percent higher and 13 percent lower ratings given to goals 3 and 5, respectively, by the experienced subjects.

The aid level data suggest that this effect produced differences in perceived utility for the aid. Comparing across aid levels indicates a tendency for the highest ratings to be given by those subjects who were exposed to the full aid condition. However, for goal 5, which was related to providing a criterion for evaluating strike timing decisions, the highest ratings were observed for the utility condition. The utility condition ratings for goal 5 were 8 to 13 percent higher than for the other goals. This finding may have been anticipated, because utility represents a fundamental comparison criterion. Across the other goals, only the ratings for goal 4 possessed any substantial variability across aid levels. Goal 4 concerned own versus enemy losses and was perceived by the subjects in the utility condition, the condition for which no loss information was available, as very far from satisfied by the aid. For goal 4, the highest ratings were obtained for the full aid condition and intermediate levels were obtained for the outcome and no uncertainty conditions.

Table 15

Results of Multiattribute Utility Analysis

Maximum Possible	(3500)	(500)	(1500)	(2500)	(1300)	(700)	(10.000)
	Utility For Goal (G)						
	G1	G2	G3	G4	G5	G6	Total Utility
Overall	2991	454	1210	1950	1036	618	8259
Experienced	2916	443	1266	2026	964	628	8243
Inexperienced	3065	464	1154	1875	1108	608	8274
Aid Levels							
Full Aid	2990	471	1269	2282	1002	630	8644
Utility	2990	467	1237	1604	1127	630	8055
Outcome	2972	431	1157	1958	980	624	8122
No Uncertainty	3011	444	1177	1958	1035	588	8213
Experienced							
Full Aid	2858	454	1200	2313	845	624	8294
Utility	2858	467	1275	1563	1137	659	7959
Outcome	2946	446	1413	2208	953	618	8584
No Uncertainty	2478	404	1175	2021	921	612	8134
Inexperienced							
Full Aid	3121	487	1337	2250	1159	636	8990
Utility	2730	467	1200	1646	1116	601	7760
Outcome	2998	417	900	1708	1007	630	7660
No Uncertainty	3022	483	1180	1896	1148	563	8292

The experience by aid level data for goal 4 suggests that, higher ratings were assigned by the subjects in the full aid, outcomes, and no uncertainty conditions than by the inexperienced subjects. The information provided by the aid may have been more meaningful to the experienced subjects.

Also showing some variability across background by aid levels were the ratings relative to goal 5. Goal 5 was concerned with the use of the ASTDA as an evaluation criterion. The experienced group rated achievement of this goal relatively low, at least in the full aid, outcomes, and no uncertainty conditions.

The background by aid level analysis also suggested some variability relative to goal 1--to assist in the derivation of the best possible strike launch time. The results indicated a rather low goal attainment evaluation by the experienced subjects in the no uncertainty condition and to a lesser extent by the inexperienced subjects in the utility condition.

After Evaluation Interview

Each subject, after completing the eight scenario problems, was interviewed relative to his impressions of the ASTDA. Information was sought about usefulness and influence of various aspects of the aid.

Input Displays

The subjects were queried about the usefulness of the input displays. They indicated their response on a five category rating scale. The mean usefulness ratings for the input displays are presented in Table 16. The input displays considered were the WAT, WAC, BFA, ORAD, and ORGF. The ORGF information was treated as a unit. The information on the desired number of blue (DNB) was presented to the subjects embedded within the BFA displays but was rated separately.

Overall, the ratings tended to vary between "3" and "4," i.e., between useful and highly useful. The highest rating, 3.84, was received by the ORGF display. The BFA and the ORAD displays were rated as 3.71 and 3.62 respectively. The lowest rating, 2.39 was assigned to the DNB (desired number of Blue). The experienced subjects generally rated the input displays to be more useful than the inexperienced subjects. This was true for every display except WAT. Possibly, the experienced subjects, by virtue of their backgrounds, were able to read more into the input displays than the inexperienced subjects.

When the data are considered across aid levels, the input information was not rated highest in the no aid condition, which had only input information available. Rather, on the average, the highest rating was observed in the utility condition which rated the input at 3.88. The input

Table 16

Mean Ratings of Input Usefulness

	WAT	WAC	EFA	DNB	ORAD	ORGF	Mean
Overall	3.42	3.02	3.71	2.39	3.62	3.84	3.33
Experienced	3.27	3.19	3.81	2.58	3.64	3.90	3.39
Inexperienced	3.56	2.86	3.60	2.20	3.60	3.77	3.27
Aid Levels							
Full Aid	3.58	3.00	2.84	2.10	3.00	3.42	2.99
Utility	4.00	3.34	4.75	2.67	4.17	4.34	3.88
Outcome	3.08	3.30	3.58	2.67	3.67	3.58	3.25
No Uncertainty	2.84	2.64	3.26	2.44	3.27	4.00	3.07
No Aid	3.58	3.25	4.17	2.09	4.00	3.83	3.49
Experienced							
Full Aid	3.33	2.83	2.67	2.20	3.00	3.50	2.92
Utility	4.17	4.00	4.83	3.00	4.67	4.67	4.22
Outcome	2.33	2.50	3.50	2.33	4.00	3.67	3.05
No Uncertainty	2.86	3.28	3.57	2.71	2.71	4.00	3.19
No Aid	3.67	3.33	4.50	2.67	3.83	3.67	3.61
Inexperienced							
Full Aid	3.83	3.17	3.00	2.00	3.00	3.33	3.06
Utility	3.83	2.67	4.67	2.33	3.66	4.00	3.53
Outcome	3.83	3.33	3.67	3.00	3.33	3.50	3.44
No Uncertainty	2.83	2.00	2.83	2.17	3.83	4.00	2.94
No Aid	3.50	3.17	3.83	1.50	4.17	4.00	3.36

information was rated lowest by subjects in the full aid and the no uncertainty condition with scores of 2.99 and 3.07, respectively. The mean ratings of the subjects assigned to the outcome condition, 3.25, was slightly higher.

These findings support contentions favoring the salience and usefulness of most of the input information provided by the ASTDA.

Outcome Displays

A parallel set of ratings was completed for the outcome displays. The overall usefulness mean, 3.95, was somewhat higher than the usefulness mean for the input information. The results, presented in Table 17, generally support the usefulness of the outcome displays.

The highest overall ratings, 4.50 and 4.14, were assigned to the ORGL and the BFL displays, respectively. The ORAL display was rated slightly lower--3.21.

The comparisons across background levels suggest that the experienced subjects tended to rate the usefulness of the outcome information lower than the inexperienced subjects, except for the ORGL display.

Examining the data across aid levels indicates that the outcome displays were rated highest in the condition that only had one source of output information (utility or outcome) available. The background by aid level analyses of Table 17 suggest that this effect was only influential on the judged usefulness of the outcome displays for the experienced subjects. Here, the effect was strong. The outcomes were rated lower in the full aid and no uncertainty conditions when both sources of output information were available than they were when only the outcome information was available. Examining the inexperienced subject data across aid levels does not indicate the same type or degree of systematic variability as observed in the experienced subject data.

Input-Outcome Interaction

The usefulness ratings by the experienced subjects can be employed to quantify further the usefulness of various displays. Usefulness may be defined:

$$VII_I = k \frac{H_I}{H_I + H_O + H_U} = k \frac{H_I}{H_{tot}}$$

Table 17

Mean Ratings of Outcome Usefulness

	BFL	ORAL	ORGL	Mean
Overall	4.14	3.21	4.50	3.95
Experienced	4.11	2.87	4.72	3.90
Inexperienced	4.17	3.21	4.50	4.00
Aid Levels				
Full Aid	3.68	3.33	4.50	3.84
Outcome	4.42	3.49	4.50	4.14
No Uncertainty	4.33	2.83	4.50	3.87
Experienced				
Full Aid	3.67	2.33	4.50	3.67
Outcome	4.50	3.30	4.83	4.21
No Uncertainty	4.16	2.50	4.83	3.83
Inexperienced				
Full Aid	3.68	3.83	4.50	4.00
Outcome	4.33	3.67	4.17	4.06
No Uncertainty	4.50	3.17	4.17	3.95

where: VH_I is the usefulness of the input information (H_I)

H_O is the information from outcome ratings

H_U is the information from utility ratings

k is a constant of proportionality ($1 \leq k \leq 5$).

Changes in VH_I as a function of changes in H_{tot} are then an expression of the fact that all information (H) is relative, and the importance (V) of one information source is indirectly determined by the availability of other sources of information. As applied to the present situation, when only input information was made available, i.e., $H_{tot} - H_I = O$:

$$VH_I = k \frac{H_I}{H_I}, \quad \text{or}$$

$$VH_I = k.$$

When other sources of information were also available and contributed to uncertainty reduction, then:

$$VH_I > k.$$

When additional sources of information acted so as to increase uncertainty, then:

$$VH_I < k.$$

The usefulness rating assigned by the subjects (3.49) to the input information (H_I), when presented in the no aid condition, was moderately high. This value would be obtained when $H_O + H_U = 0$ so that $VH_I = k$, and can be considered as the reference value against which the effects of making other information available can be compared. The usefulness value ranged between 2.92 and 3.19 when either $H_O + H_U$, or when only H_O was available along with H_I. This suggests that $H_O + H_U$ was positive, i.e., contributed to uncertainty reduction. VH_I rose sharply to 4.22 when only H_U was made available, suggesting that $H_O + H_U$ was negative. The utilities without the outcome information did not reduce uncertainty and possibly increased it.

The usefulness of outcome displays is given by:

$$VH_O = k \frac{H_O}{H_I + H_O + H_U} = \frac{H_O}{H_{tot}}.$$

Since H_I was constant across all conditions, the only interest is in the effect of incrementally providing utility displays (H_U) for experienced subjects with respect to H_O. When only outcome information was made available, VH_O = 4.21. Additionally providing H_U reduced VH_O to an average value of 3.75. This decrease strongly suggests that H_U increased H_{tot}, by reducing uncertainty for experienced subjects. These findings allow some additional insight into and refinement of our understanding of expected utility (EU) in particular and of the information presented in general. Apparently, EU (or H_U) only presented in conjunction with input information (H_I) had little, if any, beneficial effect, but, when presented along with the outcomes (H_O) it did have a beneficial effect. This suggests a simple notion that providing increased amounts of information (ΔH_{tot}) will necessarily increase the value of the information (ΔVH_{tot}) for the user is misleading. Again, as indicated by the variance analyses, more is not necessarily better.

Influence of Outcome Displays

The subjects were also asked to rate the influence of the output displays on their strike timing decisions. Again, five category scales ranging from "no influence" to "very much influence" were employed. The data derived from these questions are presented in Table 18. For the overall analysis, the highest influence rating of 4.47 was assigned to the ORGL followed by similar ratings of 4.02 and 3.97 for the EU and the BFL, respectively. The lowest rating, 3.08, was received by the ORAL display.

The data for the experienced subjects suggest that they were most influenced by the ORGL outcome while the inexperienced subjects were most influenced, on the average, by the EU information.

The comparison across backgrounds by aid levels suggests that for the experienced subjects the highest ratings came from the utility and outcome conditions with averages of 4.17 and 4.19, respectively. With almost perfect consistency, the outcome displays were rated lower in influence in the full aid and no uncertainty conditions with averages of 3.67 and 3.81, respectively. That is when both the outcome and the utility displays were available, they were rated lower than when either one set or the other was available.

The distribution of the influence ratings by the inexperienced subjects showed few systematic differences across aid levels. This finding parallels that already reported in the usefulness ratings.

In keeping with the reasoning and notation above:

$$VH_i = k \frac{H_i}{H_{tot}}$$

Table 18

Mean Ratings of Output Influence

	BFL	ORAL	ORGL	EU	Mean
Overall	3.97	3.08	4.47	4.02	3.93
Experienced	3.83	2.94	4.72	3.78	3.82
Inexperienced	4.11	3.23	4.22	4.28	3.96
Aid Levels					
Full Aid	3.75	3.08	4.08	3.75	3.67
Utility				4.17	4.17
Outcome	4.58	3.25	4.75		4.19
No Uncertainty	3.58	2.93	4.58	4.17	3.81
Experienced					
Full Aid	3.50	3.00	4.33	3.00	3.46
Utilities				4.33	4.33
Outcome	4.67	3.17	4.83		4.22
No Uncertainty	3.33	2.66	5.00	4.00	3.75
Inexperienced					
Full Aid	4.00	3.16	3.83	4.50	3.87
Utility				4.00	4.00
Outcome	4.50	3.33	4.67		4.17
No Uncertainty	3.83	3.20	4.16	4.33	3.88

where H_i represents information from display i. Therefore, any factor that increases $H_{tot} - H_i$ will decrease VH_i. This condition would be maximal when, both, outcome and utility displays were available and produced positive modifying effects and a relatively low VH_i. When either utility or outcomes were the only output information available, $VH_i = 4.28$ on the average, suggesting that $H_{tot} - H_i$ was comparatively small. When, both, outcome and utility displays were concurrently available, $VH_i = 3.60$ on the average, suggesting that $H_{tot} - H_i$ was large. Therefore, it could be argued that the influence of information on the strike timing decisions, like usefulness, could be considered relativistic, being inversely related to other modifying information concurrently available. Note that these effects were only consistently observed in the data for the experienced subjects. The finding seems reasonable when one considers the fact tha' the effect is essentially a measure of the relative sensitivity of decision makers to information from various sources. The extent to which any set of stimuli is informative depends on the background and experience of the decision maker with respect to the meaning of the information. Sensitivity, therefore, is a function of experience. The experienced subjects were apparently more sensitive to the information because they have had more experience with strike launch situations and a consequent better understanding of possible effects of each set of information.

Discussion of Ratings

The ratings provide some important insights about the ASTDA. The experienced subjects apparently considered the input information to be more useful than the inexperienced subjects and generally the output information was less important to the experienced subjects than to the inexperienced subjects. However, both groups rated the outcome information as the more important. Consistently the ORAL and BFL information was indicated to be the most important to the experienced subjects. Even with this consistency, the experienced subject data indicated that they were very sensitive to the type of information available. This finding provides some explanation for the previously reported result that our expert panel selected preemptive strike times for two problems. Evidently the panel members, as compared with our subjects, were differentially sensitive to the information provided.

Other Opinions

A set of open ended questions was included in the interview to allow for the derivation of qualitative information about the aid. The information from the experienced, fully aided subjects will be used to draw a picture of the aid as they saw it. The information from the other subjects and aid level conditions will then be discussed in so fas as it adds detail or new elements to the discussion.

Generally, the fully aided, experienced subjects perceived the aid as "pretty helpful" and indicated that it "could be a good tactical decision aid." These statements tended to be qualified on two counts: (1) that the aid is only as good as the input information, and (2) that the algorithms built into the aid are reasonable. The concern about input information was rather consistently stated. There seemed to be a pervasive feeling that information supplied by weather and by intelligence officers tends to be less than fully reliable. Because a considerable part of the ASTDA's information is based on these sources (WAT, WAC, ORAD, and ORGF), and because the ASTDA's outputs are derived from these sources, it was generally indicated that for the aid to be useful, this information had to be accurate. Concerns about the algorithms were fewer and mostly related to the utility measure. There was a tendency to question the expected utilities as being "too inclusive" or "too general."

The subjects rarely questioned the validity of the loss information (BFL, ORAL, ORGL). In fact when the input information was divergent, the subjects said that they tended to base their decisions on the loss information. Specifically, they said that they attempted to balance the information on their own losses and the destruction of enemy targets. Also, some subjects said that they compared their decision to the EU information and, if there was an important discrepancy, they reconsidered their original decision. However, when the input information was convergent, and alternatives were rather obvious, the decisions tended to be based on input information. Therefore, the usefulness of the output information tended to be somewhat proportional to the divergence of the input information.

Other interview questions examined specific aspects of the aid, e.g., the display formats, advantages of color, etc. When asked if the tabular or the graphic information presentation was more useful, the subjects overwhelmingly chose the tabular format. They generally thought that the graphs were difficult to read. The importance of color for the graphs was also rated rather low--helping as a discriminant but no more.

When asked about the usefulness of the averages and ranges displayed on both the tables and graphs, the responses were more variable. About half of the subjects said that the averages were more helpful while the other half thought that both the averages and ranges were helpful. Subjects who preferred the averages wanted to see things at a glance or to obtain a ready indicator while those who preferred both said that the averages were deceptive and that the range information presented a better picture of what to expect.

When questioned as to whether or not the ASTDA helped to have more confidence in decisions, the overwhelmingly answer was affirmative. The reasons given were that the organization of the information tended to

focus thinking on a few particulars, to indicate possible trends, and possibly reduce human error. The one person that answered negatively qualified his statement by noting that while not increasing his confidence, the aid certainly provided a faster means for deriving a decision.

Similar affirmative answers were also obtained when the experienced subjects were asked if ASTDA aided decisions were better decisions than non ASTDA aided decisions, and if they would feel comfortable using ASTDA during actual combat conditions. However, laced through their positive responses to ASTDA were again the qualifications that they would be "confident," "comfortable," and "make better decisions" only if the information entered into the system was accurate.

The subjects also were asked about the additional information which the ASTDA should supply and if they had any further comments. The responses ranged over a wide area. Several experienced subjects indicated that the aid should include a psychological readiness of pilots factor which could interact with other factors and affect outcomes. Such behavioral modeling is well within the current state-of-the art. It was also suggested that the aid does not consider a range of relevant strategies, e.g., with a low-probability of good visibility at the target, the cloud cover could be used strategically (sending two strikes, one above and one below the cloud cover).

Similarly, it was suggested that the aid should consider various mixes of armament and ordinance. Others thought that the aid did not address some very important points, e.g., search air rescue, refueling time after launch, rendezvous times and places, as well as some minor points, e.g., aborts of the mission not due to enemy actions.

One area of recurring concern had to do with the graphic displays. It was suggested repeatedly that relevant graphs should be either superimposed or presented simultaneously on a split screen, or nomographically, or in some combination which would simplify comparisons. It was also suggested that in the outcome displays (BFL, ORAL, ORGL), the lost or destroyed units should be weighted and summed. In addition, it was suggested frequently that the information on orange ground target availability and destruction should be separated from the information on orange ground defenses likely to be encountered and destroyed.

The information obtained from the experienced subjects exposed to the other aid conditions supplements the prior considerations. One aspect which seems rather relevant concerns the expressed need of the utility condition subjects for specific loss information, and by the outcome condition subjects for a general measure of trade-off or utility. This finding supports the need for such information as included in the aid. The discussion

of influence ratings (in section on Influence of Outcome Displays) also indicated the augmental/supplemental nature of the two types of display. We also note our prior finding of no statistically significant differences between the experimental conditions in which either one or the other of these types of display was available but the statistically significant superiority of the fully aided over the partially aided conditions. Subjects from these conditions also tended to express the need for accurate algorithms and, to a lesser extent, accurate input information.

The subjects in the utility condition generally tended to report having based their decisions on input information but they also tended to be sensitive to differences in utilities across times, suggesting the need for some measure of significance. The answers and suggestions received from the no uncertainty condition subjects indicated little appreciation for the mission uncertainty indicators. None of the subjects reported the need for such measures.

There were also some questions as to the accuracy of the expected utility output of the aid. As was previously stated, the possibility that the utilities were too inclusive was mentioned. It was also said that the perceived value of the expected utility might be enhanced if the user was informed about what units were included in calculating the utilities as well as specifying the value of each included unit.

The reasonableness of any specific item of outcome information was not questioned. This obtains in spite of the stated need for reasonably accurate algorithms. This result may have been related to the subjects' attitude when asked if they would use the aid under live combat conditions. The responses were generally that they would certainly use it as an additional set of information that would have to be considered when making a strike timing decision. When asked for the reason, the usual answer was that the ASTDA stands head and shoulders above the competition (because there is nothing to compare it to).

IV. DISCUSSION AND SUMMARY

What then may be said about the value of the ASTDA in particular and about the implications of the present work for decision aids in general?

The results of the present work certainly seem to support contentions favoring the value of the ASTDA. The variance analyses consistently indicated statistically significant differences between the aid levels investigated. And, where differences were found, they were between the unaided condition and some level of aiding. On the other hand, the results suggested that although some level of aiding helps, more is not necessarily better. It seems that the aid's input displays acted as an information organizational tool. They set the information that the user wanted to consider into perspective and into meaningful relationships. The user employed only parts of the information. Similarly, multiple and vast arrays of output seemed to add little. Once the input was organized, the user cared little about multiple outputs which he could not mentally synthesize into an integrated whole. Accordingly, he selected the output(s) or major meaningfulness to him and rested with it (them). While the number of input and output displays which the user can manage is not known, the number is certainly fewer than the number provided by the ASTDA. Note that the subjects in the evaluation indirectly voiced this same thought when they suggested, during the interview, that split screen displays and nomographs for relating displayed information would be helpful.

One limitation of the aid was its failure to accommodate situational variables which may bias, and possibly override, the data produced by the aid. Specifically, the results from the aid did not square with the conclusions of our panel on the matter of preemptive strikes. The aid did not consider such data biasing situations. It seems that such overriding variables or contingency conditions should be taken into account during the design of any decision aid. Otherwise, the aid will fail to provide full realism with the result that its acceptability may suffer.

The interaction effects noted by the various analyses are reasonable but add complexity to the aid design problem. Output information was significant for "hard" problems but not for "easy" ones. For "easy" problems the input displays were most meaningful. Moreover, there was some evidence that the output displays achieve much of their value by virtue of their ability to sensitize the user to the input information. Seeing the projected outcome forces the user to ask, in a sense, what could cause that? He may then reconsider an earlier decision or generate additional hypotheses for investigation through the use of the aid. Both the multiattribute utility and the expert panel data suggested that such nuances were more readily perceived by the experienced than by the inexperienced subjects--as might have been anticipated.

Certainly, the ASTDA has achieved its goals, as defined within the multiattribute utility analysis, to a considerable extent. The multiattribute utility of the aid across the six goals considered was 83 percent of the total possible utility. Utility relative to individual goals ranged from 79 percent to 91 percent of the total possible. These values seem quite high--especially since the ASTDA, as tested, was not necessarily in its final form.

From the point of view of the relative merit of the aid, our analysis indicated an increase in decision validity by a factor of five when unaided are compared with aided decisions.

The after experiment interview indicated a number of areas for attention within the ASTDA itself. These generally included uncluttering and integrating the various displays. Moreover, according to our subjects, the color feature and the color graphics added little. Some cost savings might be implemented by eliminating these factors. Certainly, the operational acceptance of the ASTDA will depend on the availability of reliable input information and on the faith of the user in the internal algorithms. We are in no position to judge either the reliability of the input information demanded by the ASTDA or the validity of its algorithms. However, any ultimate, fleet user's orientation should address and make information available about these issues.

None-the-less, the subjects found the ASTDA to provide useful information which influenced their strike timing decisions. The experienced Navy flight officer subjects said that they believed ASTDA aided decisions to be superior to nonaided decisions and that they would feel comfortable using the ASTDA during actual combat.

Implications for Future Evaluations

The present work also provided a number of methodological insights which should be considered in any evaluation of a decision aid which is conducted in the future.

First, the criterion problem remains open. The present study attempted to come to grips with the criterion problem, at least partially, by employing two criteria. The use of multiple criteria has been advocated in other fields (e.g., test development). But, such an approach does not provide an answer to criterion reliability problems. We have no data relative to the reliability of the launch time judgments of our criterion panel and the question of whether or not our panel would provide similar results on a retest remains open. Moreover, the panel rankings did not agree entirely with our second criterion, the utility rankings produced by the aid. The reasons for this were given earlier. However, if two criteria agree only moderately or disagree, how can one expect to obtain significant

validity for the aid against each of the criteria taken separately? Surely, if the aid agrees with one criterion it will disagree with the other. If a mathematical solution to a problem disagrees with the best judgment of management, which course of action is to be preferred? Management will want to know the assumptions of the mathematical solution and methods in such a case and, once aware of these, management may or may not accept the mathematical solution. Additionally, in the present case, the total situation becomes more circular because the mathematical solution (utility) was itself a part of the aid and was available to the subjects in three of the four aided conditions (Exhibit II). Accordingly, employing this criterion presents the situation of assessing the aid against its own output. Yet, surprisingly, our data did not indicate a strong reliance by the subjects on the expected utility output of the aid. Possibly they did not understand or trust the utility construct. In sum, although we employed multiple criteria and continue to advocate such an approach in aid evaluation, such an approach does not compensate for criterion weakness.

Second, any aid evaluation will depend on the state of development of an aid at the time at which it is evaluated. Evaluating too early may result in an injustice to an aid because the aid developers may not have had sufficient opportunity to refine their design. Evaluating too late may allow errors to go unrecognized until it is too late to do anything about them. Accordingly, as suggested by Figure 1, aid evaluation may need to be viewed against a continuum rather than as a process to be carried out at a specific point in time. And, any evaluative results are pertinent only to the state of the aid at the time at which it was evaluated.

When one is involved with laboratory experiments, he must be content with intermediate criteria. Such intermediate criteria are more often than not based on matters of practicality rather than true relevence to the ultimate criterion. Similarly, criterion sensitivity becomes an issue. Our failure to find differences between some of the levels of aiding may be a function of lack of criterion sensitivity rather than any failure of the aid. The subjects may have worked harder in one condition as compared with another. But, this was not measured by the criteria employed.

In retrospect, it becomes apparent that the use of the inexperienced group as subjects may not have been warranted. While the information/data from such subjects is of theoretic interest, such theoretic excursions are costly. After all, in actual practice, one can anticipate that decisions, such as those with which we were concerned in the present work, will be made by experienced persons.

The interactive and the moderator effects which were evident in several of the analyses point up the fact, known at the outset, that aid development and aid evaluation are not easy ways to pass one's time. Evaluations must be carefully designed to allow for the identification of such effects, if present. Barren research designs will miss such nuances. And, these may be more important than "main" effects.

Finally, we note that a number of our analyses were based on correlational methods. Correlation implies strength of association--not causality. Because correlation is a fundamental tool of the behavioral sciences, a number of techniques have been developed which allow one to go beyond mere statements of relationship on the basis of correlation and to derive statements of causality. Most, if not all, of these methods are based on structural equation models, and the models have been variously referred to as simultaneous equation systems, linear causal analysis, path analysis, structural equation models, dependence analysis, cross-lagged correlation, and the like. The end result is statements of cause and effect and because the end result represents a causal link rather than a measure of association, the structural results do not coincide, in general, with coefficients of regression among observed variables.

Background material on structural equation models may be found in Heise (1975), Duncan (1975), and Goldberger (1972). Two volumes by Blalock (1971, 1974) contain several papers dealing with basic issues and problems at an elementary level. At a more advanced level, two recent volumes, Goldberger and Duncan (1973) and Aigner and Goldberger (1977) cover several issues, problems, and applications. Bielby and Hauser (1977) gave an excellent review of the sociological literature on structural equation models.

One of the more recent and sophisticated techniques which relies on linear structural equations was described by Joreskog and Sorbon (1978)--analysis of linear structural relationships by the method of maximum likelihood (LISREL). The LISREL model was designed to handle models with latent variables, measurement errors, and reciprocal causation. The model seeks to establish whether or not a causal relationship exists among a set of latent variables--some of which are designated as independent variables while others are designated as dependent variables. The procedure also possesses the important advantage that it requires measurement at only one point in time--a distinct practical advantage. The structural equation model specifies the causal relationship among the latent variables and the amount of unexplained variance. Joreskog and Sorbon (1978) published and made available a general LISREL computer program for IBM systems for deriving the required structural equations. It would seem that such causal relationships should be explored in future evaluations of the sort reported here.

Original Hypotheses

In Chapter I, five hypothese were established to provide a basis for the present evaluation. Each of these is now discussed. Hypothesis 1, more effective strike time decisions can be made with the aid then without the aid, seems to be quite strongly supported by the data. If effectiveness is defined

as how closely the subjects' decisions approximated the predictions of the aid, then ASTDAs effectiveness was certainly demonstrated by the rather consistent differences observed between the full aid and the no aid conditions. Arguments supporting the apparent effectiveness of the aid seem to be further supported by the findings that, with only parts of the information provided by the aid, the quality of the decisions did not suffer. The lack of differential effectiveness across the aided conditions suggests that Hypothesis 5, decision effectiveness will vary systematically as the characteristics of the aid are varied, cannot be supported or accepted. Whether or not the aid's design represents a case of overkill is difficult to say. Perhaps the obvious has been restated, i.e., the human can only manage a limited amount of information at one time. The effectiveness of the ASTDA is probably at least a partial function of its sensitizing, feedback mechanisms.

The algorithmic logic of the ASTDA is not obvious to the user--especially how nontransparent changes in particular variables might influence outcomes. It might make sense to provide the user with this type of information. This information might be particularly important if the aid were accepted for use in the fleet and might be an integral part of the training of: (1) personnel who use the aid, and (2) those who review any decisions derived from the use of the aid. In addition, such information would probably be useful in allowing a user to get a "feel" for the significance of the various aspects of the aid.

Hypothesis 2, users will perceive the aid to possess value (utility), was also supported by the data from a variety of sources. In terms of overall assessment, both the data from the perceived utility in relation to the goals and the data from the interview suggest that the ASTDA possesses qualified value. The qualifications, as already stated, concern the input information and the aid's algorithms.

Hypothesis 3, effectiveness and perceived value will not vary as a function of the user's experience or problem difficulty, cannot be unequivocally accepted. The effectiveness of the decisions did not vary with experience but did vary with difficulty. However, the first order interaction indicated that most of the effects of difficulty were produced in the no aid condition. Perceived value of the aid, depending on the measure, fluctuated only in minor ways with experience. This was certainly true of the perceived utility in relation to the ASTDA goals. However, the subjects' assessments during the interview suggest that perceived utility may increase as a function of difficulty or divergence of the information.

Support for the Hypothesis 4, the strike timing decision aid possesses criterion related validity, varies in accordance with the set of assumptions one accepts. Using the expert panel's judgments as the validity criterion

and assuming that the aid should rank order alternate strike times in a manner congruent with those of the criterion, suggests moderate criterion related validity for the aid. If it is assumed that the ASTDAs validity should be assessed only within the limited area that it addresses, then the agreement level increases. That is, the validity rises if the possibility of preemptive strikes is not considered. If it is further assumed that the only important decision facing a task force flight operations officer is the one time to launch an air strike rather than a ranking of times, then the apparent validity based on agreement rises considerably.

REFERENCES

Aigner, D. J. and Goldberger, A. S. Latent variables in socioleconomic models. Amsterdam: North Holland Publishing Company, 1977, 7.

Analytics, Inc. Measuring the Performance of Operational Decision Aids. ONR Report 1161-B. Willow Grove, Pa: Analytics, Inc., April, 1976.

Bieler, W. T. and Hauser, R. M. Structural Equation Models. Annual Review of Sociology, 1977, 3.

Blalock, H. M. Cau-al models in the social services. Chicago: Aldine-Asterton, 1971.

Blalock, H. M. Measurement in the social services. Chicago: Aldine Publishing Company 1974.

Brown, R. V. Research on Decision-Analytic Technology, ONR Report PR 78-11-25. McLean, Virginia: Decision and Design, Inc., May, 1978.

Chapanis, A. Man, Machines, and Models. American Psychologist, 1961, 16, 113-131.

Christal, R. E. Selecting a harem-and other applications of the policy capturing model. Journal of Experimental Education, 1968, 4, 35-41.

Duncan, O. D. Introduction to structural equation models. New York: Academic Press, 1975.

Edwards, W. Social Utility paper presented at the Symposium on Design and Risk Analysis, Annapolis, Md.: American Society for Engineering Education and American Institute of Industrial Engineers, 1971.

Glenn, Floyd. ASTDA User's Guide. ONR Report 1344-A. Willow Grove, Pa.: Analytics, Inc., 1978.

Goldberger, A. S. Structural equation methods in the social sciences, Econometrica, 1972, 40, 979-1001.

Goldberger, A. S. and Duncan, O. D. Structural equation models in the social services. New York: Seminar Press, 1973.

Heise, D. R. Causal analysis. New York: Wiley, 1975.

Joreskog, K. G. and Sorbon, D. LISREL: Analysis of linear structural relationships by the method of maximum likelihood. Chicago: National Educational Resources, 1978.

Lucas, S. L. and Ruff, J. A. An Investigation of Operational Decision Aids. ONR Report 312. Arlington, Virginia: Systems Planning Corporation, July, 1977.

Payne, J. R. and Rowney, J. V. ONRODA Warfare Scenario (SRI Research Memorandum NWRC-RM-83). Menlo Park, CA: Stanford Research Institute, June, 1975.

Rowney, J. V. Amphibious Warfare Scenario (SRI Research Memorandum NWRC-RM-86). Menlo Park, CA: Stanford Research Institute, October, 1975.

Siegel, A. I. and Madden, E. G. Annual Report Decision Aid Evaluation Program. Wayne, Pa.: Applied Psychological Services, Inc., June, 1979.

Siegel, A. I. An Experimental Design for Evaluating the "Strike Timing Aid." Prepared for ONR. Wayne, Pa.: Applied Psychological Services, Inc., 1978.

Sinaiko, H. W. Operational Decision Aids: A Program of Applied Research for Naval Command and Control Systems. ONR Report TR-5. Washington, D.C.: Manpower Research and Advisory Services, Smithsonian Institution, July, 1977.

Appendix A

Postevaluation Interview Form

ASTDA POST EXPERIMENTAL DEBRIEFING FORM

Subject Name ______________________ Date ____________
Subject ID Number ______________________
Treatment Condition ______________________
Experienced ____Yes ____No
Interviewer ______________________

As the final part of experiment, I have around 20 questions to ask you. These questions concern your opinions about the various features of ASTDA and their value. Okay.

1. Please tell me, in general terms, about how useful, if at all useful, the ASTDA was to you for coming to a direct strike timing decision?

____
____
____
____
____

2. How did the ASTDA help you most for deriving a strike timing decision? How so? Least? How so?
   Most help: ____
   ____
   How so: ____
   ____
   Least help: ____
   ____
   How so: ____
   ____

3. What aspects of the ASTDA were most confusing to you? How so?
   Most confusing aspects: ____
   ____
   ____
   ____
   How so: ____
   ____
   ____
   ____

4. Would you say that ASTDA was (a) more useful in deriving a strike timing decision for some problems than for others or (b) was ASTDA equally useful for all problems? How so?

_____ (a) _____ (b)

How so: ____

____

____

____

____

____

5. With this usefulness rating scale (show card with scale) rate the usefulness of the information for deriving your strike time decisions. What is your rating for the information on:

	Rating				
Readiness and Weather Report Time	1	2	3	4	5
Air Strike Mission Structure	1	2	3	4	5
Weather at Target	1	2	3	4	5
Weather at Carrier	1	2	3	4	5
Blue Force Readiness	1	2	3	4	5
Desired Number of Blue Aircraft	1	2	3	4	5
Orange Air Defense	1	2	3	4	5
Orange Ground Force	1	2	3	4	5

6. (For treatment conditions 1, 2, and 4 only)
How much influence did the utility outcome values have on your stike timing decisions? Would you say:

	Rating
no influence at all	1
very little influence	2
some influence	3
much influence	4
very much influence	5

7. (For treatment conditions 1, 3, and 4 only)
With this rating scale (show card with usefulness scale), rate the usefulness of the following combat loss information:

	Rating				
1. blue force air losses versus strike time	1	2	3	4	5
2. orange air losses versus strike time	1	2	3	4	5
3. orange ground losses versus strike time	1	2	3	4	5

8. (For treatment conditions 1, 3, and 4 only)
Using this scale (show card with influence scale), rate how much influence each combat loss information aspect had on your strike timing decisions:

	Rating				
1. blue force air losses versus strike time	1	2	3	4	5
2. orange air losses versus strike time	1	2	3	4	5
3. orange ground losses versus strike time	1	2	3	4	5

9. Tell me the order of usefulness--from most to least useful--of each of these (show cards with types of information) in helping you decide on the time to launch an air strike. Which was first most useful, which was second, and so on until you have ordered the information from first to last?

(For treatment condition(s) only)

		Rank
ALL	Readiness and Weather Report Time	______
ALL	Air Strike Mission Structure	______
ALL	Weather at Target	______
ALL	Weather at Carrier	______
ALL	Blue Force Readiness	______
ALL	Desired Number of Blue Aircraft	______
ALL	Orange Air Defense	______
ALL	Orange Ground Force	______
(1, 2, and 4 only)	Utility Outcome Values	______
(1, 3, and 4 only)	Blue Force Air Losses	______
(1, 3, and 4 only)	Orange Air Losses versus Strike Time	______
(1, 3, and 4 only)	Orange Ground Losses versus Strike Time	______

10. Was the tabular or graphic information more useful, or were both the tables and graphs equally useful? Why do you think so?

1. tables more useful
2. graphs more useful
3. both equally useful

Why? ____
____
____
____
____
____
____

11. (For treatment conditions 1, 2, and 3 only)
Were the averages or the ranges provided on the tables more useful, or were the averages and ranges both equally useful? How so?

1. averages more useful
2. ranges more useful
3. both equally useful

How so? ____

____

____

____

____

12. Using the usefulness scale (show card with scale), rate how useful, if at all, the coloring of the graphs.

Rating
1 2 3 4 5

Why did you rate this so? ____________________________

____

____

____

____

13. (For treatment condition 1, 2, and 3 only)
Using the usefulness scale (show card with scale), rate the usefulness of the delta-biased uncertainty bands.

Rating
1 2 3 4 5

Why did you rate this so? ____________________________

____

____

____

____

14. What, if anything, was confusing about the tables? Which tables were confusing? How so?
What? ____

____

____

Table Name(s) ______________________________________

____

____

____

How so? ____

____

____

____

15. What, if anything, was confusing about the graphs? Which graphs were confusing and how so?

What? ____

____

____

____

Graph Name(s) ____

____

____

____

How so? ____

____

____

____

16. What additional information do you think the ASTDA's tables and graphs should include? Why do you think so?

Tables ____

____

____

____

Graphs ____

____

____

____

Why? ____

____

____

17. What information provided or made available on the ASTDA's tables and graphs do you think should be deleted? Why do you think so?

Tables ____

____

____

____

Graphs ____

____

____

____

Why? ____

____

____

____

18. Do you think the use of the ASTDA helped you to have any more confidence in your strike timing decisions than you would have had if you did not use the ASTDA? How so?
_____ Yes _____ No
How so? ____
____
____
____

19. Do you think the ASTDA aided decisions are better decisions than non ASTDA aided decisions. (Why do you say yes? Why do you say no?)
_____ Yes _____ No
Explain ____
____
____
____

20. (For experienced group only)
Would you feel "comfortable" in using the ASTDA or some variant of it under live and real combat conditions? Do you think other air operations officers would? Why?

You: _____ Yes _____ No
Others: _____ Yes _____ No

Why? ____
____
____
____

21. (For experienced group only)
Would you be surprised if the actual blue and orange combat losses during a strike mission were far worse or far better than the losses predicted by the ASTDA? Why do you think so?
____ Yes ____ No
Why? ____
____
____
____

22. Do you have any other comments?
_____ Yes _____ No
Why? ____
____
____
____

THANK YOU FOR YOUR COOPERATION

You have just read the goals of ASTDA. Now, I want you to indicate below how closely each goal was achieved by the decision aid. For each goal use a 0% to a 100% rating: 0% means the goal was not at all attained, and 100% means the goal was completely achieved. Use whatever % you feel expresses ASTDAs achievement of the goal.

Goal 1________

2________

3________

4________

5________

6________

Appendix B

Zero Order Correlation Matrices from Which the Multiple Correlation Coefficients and Regression Equations Were Computed

HARD PROBLEMS

	VAR001	VAR002	VAR003	VAR004	VAR005	VAR006	VAR007	VAR008	VAR009	VAR010	VAR011	VAR012
VAR001	1.00000	-0.15999	0.15863	0.15504	-0.00072	0.31087	0.30210	0.44265	0.00431	-0.06935	0.02761	0.40300
VAR002	-0.15999	1.00000	-0.04430	0.12490	-0.60736	0.00132	0.20890	-0.62576	0.07901	-0.00839	0.04095	0.43306
VAR003	0.15863	-0.04430	1.00000	0.20700	0.42164	-0.34937	-0.39770	0.65137	-0.06312	0.13386	-0.78994	-0.33623
VAR004	0.15504	0.12490	0.20700	1.00000	-0.46672	-0.11496	-0.09630	0.30574	-0.02524	-0.37994	0.29914	0.49090
VAR005	-0.00072	-0.60736	0.42164	-0.46672	1.00000	0.10845	-0.27373	0.43715	-0.34121	0.44101	-0.77430	-0.30667
VAR006	0.31087	0.00132	-0.34937	-0.11496	0.10845	1.00000	0.67101	0.34320	0.40770	-0.10835	0.20200	0.61960
VAR007	0.30210	0.20890	-0.39770	-0.09630	-0.27373	0.67101	1.00000	0.07225	0.40565	-0.30725	0.35655	0.46275
VAR008	0.44265	-0.62576	0.65137	0.30574	0.43715	0.34320	0.07225	1.00000	-0.40151	-0.09364	-0.45056	0.27046
VAR009	0.00431	0.07901	-0.06312	-0.02524	-0.34121	0.40770	0.40565	-0.40151	1.00000	0.04678	0.79882	0.61513
VAR010	-0.06935	-0.00839	0.13386	-0.37994	0.44101	-0.10835	-0.30725	-0.09364	0.04678	1.00000	-0.23341	-0.16069
VAR011	0.02761	0.04049	-0.78994	0.29914	-0.77430	0.20200	0.35655	-0.45056	0.79882	-0.23341	1.00000	0.66211
VAR012	0.40300	0.43306	-0.33623	0.49090	-0.30667	0.61960	0.46275	0.27046	0.61513	-0.16069	0.66211	1.00000

EASY PROBLEMS

	VAR001	VAR002	VAR003	VAR004	VAR005	VAR006	VAR007	VAR008	VAR009	VAR010	VAR011	VAR012
VAR001	1.00000	0.00210	0.01679	-0.03767	0.05708	0.67236	0.49846	0.46476	0.36040	0.28061	0.37613	0.44959
VAR002	0.00210	1.00000	-0.55359	0.25176	0.12197	0.24796	0.13128	-0.14203	0.76213	0.43492	0.58508	0.36984
VAR003	0.01679	-0.55359	1.00000	-0.09499	0.57672	-0.17581	-0.11711	0.24905	-0.55102	0.25191	-0.04042	0.18627
VAR004	-0.03767	0.25176	-0.09499	1.00000	0.11744	0.31052	0.45010	0.63925	0.36155	0.14125	0.57191	0.55211
VAR005	0.05708	0.12197	0.57672	0.11744	1.00000	-0.03033	-0.20917	0.18190	-0.11278	0.42259	0.30869	0.34376
VAR006	0.67236	0.24796	-0.17581	0.31052	-0.03033	1.00000	0.03179	0.73223	0.72502	0.47032	0.71265	0.76728
VAR007	0.49846	0.13128	-0.11711	0.45010	-0.20917	0.03179	1.00000	0.69836	0.62309	0.42619	0.62131	0.69263
VAR008	0.46476	-0.14203	0.24905	0.63925	0.18190	0.73223	0.69836	1.00000	0.30203	0.31725	0.62326	0.76432
VAR009	0.36040	0.76213	-0.55102	0.36155	-0.11278	0.72502	0.62309	0.30203	1.00000	0.54119	0.78063	0.65733
VAR010	0.28061	0.43492	0.25191	0.14125	0.42259	0.47032	0.42619	0.31725	0.54119	1.00000	0.78782	0.78417
VAR011	0.37613	0.58508	-0.04042	0.57191	0.30869	0.71265	0.62131	0.62326	0.78063	0.78782	1.00000	0.94977
VAR012	0.44959	0.36984	0.18627	0.55211	0.34376	0.76728	0.69263	0.76432	0.65733	0.78417	0.94977	1.00000

INEXPERIENCED SUBJECTS

VAR001	1.00000	-0.06325	0.06517	0.00434	0.04245	0.52329	0.38062	0.43326	0.21957	0.14706	0.20214	0.38418
VAR002	-0.06325	1.00000	-0.60062	0.36172	-0.13299	0.15779	0.08447	-0.27269	0.77686	0.26373	0.71380	0.39215
VAR003	0.06517	-0.60062	1.00000	0.01649	0.50088	-0.24011	-0.21999	0.36224	-0.64448	0.22272	-0.21507	0.11393
VAR004	0.00434	0.26172	0.01649	1.00000	-0.06645	0.14583	0.19632	0.57871	0.23604	0.05830	0.55323	0.57725
VAR005	0.00215	-0.13299	0.50088	-0.06645	1.00000	0.01667	-0.16720	0.20475	-0.17330	0.40129	-0.08566	0.15056
VAR006	0.52329	0.15779	-0.24011	0.14583	0.01667	1.00000	0.77400	0.60175	0.64972	0.27385	0.45540	0.66563
VAR007	0.38062	0.08447	-0.21999	0.19632	-0.16720	0.77400	1.00000	0.48387	0.55417	0.20219	0.36312	0.52843
VAR008	0.43326	-0.27269	0.36224	0.57871	0.20475	0.60175	0.48387	1.00000	0.09091	0.19797	0.29057	0.66319
VAR009	0.21957	0.77686	-0.64408	0.23684	-0.17330	0.64972	0.55417	0.09091	1.00000	0.38053	0.71575	0.59991
VAR010	0.14786	0.26373	0.22272	0.05830	0.40129	0.27345	0.20219	0.19797	0.38053	1.00000	0.43382	0.57835
VAR011	0.20214	0.71380	-0.21507	0.55323	-0.08566	0.45540	0.36312	0.29057	0.71575	0.43382	1.00000	0.84883
VAR012	0.38418	0.39219	0.11393	0.57725	0.15056	0.66543	0.52843	0.66319	0.59991	0.57835	0.84883	1.00000

EXPERIENCED SUBJECTS

VAR002	-0.07458	1.00000	-0.61151	0.25582	-0.12907	0.17792	0.00745	-0.24679	0.78138	0.29328	0.70540	0.38730
VAR003	0.07267	-0.61151	1.00000	0.02651	0.49183	-0.23397	-0.20408	0.37533	-0.64305	0.22933	-0.21380	0.11775
VAR004	0.02737	0.25582	0.02651	1.00000	-0.04873	0.16939	0.20380	0.58108	0.22898	0.05954	0.55837	0.57808
VAR005	0.02679	-0.12907	0.49183	-0.04873	1.00000	0.05181	-0.14092	0.23290	-0.19205	0.39907	-0.07803	0.17659
VAR006	0.51302	0.17792	-0.23397	0.16939	0.05181	1.00000	0.77603	0.60379	0.65625	0.20132	0.47297	0.67790
VAR007	0.39241	0.00745	-0.20408	0.20380	-0.14092	0.77603	1.00000	0.49232	0.54604	0.19125	0.37185	0.53712
VAR008	0.44602	-0.24679	0.37533	0.58108	0.23290	0.60379	0.49232	1.00000	0.09204	0.21293	0.38611	0.67582
VAR009	0.28452	0.78138	-0.64305	0.22898	-0.19205	0.65625	0.54604	0.09204	1.00000	0.37814	0.71588	0.59893
VAR010	0.13123	0.29328	0.22933	0.05954	0.39907	0.20132	0.19125	0.21293	0.37814	1.00000	0.43346	0.57827
VAR011	0.21287	0.70540	-0.21380	0.55837	-0.07803	0.47297	0.37185	0.38611	0.71588	0.43346	1.00000	0.85071
VAR012	0.38710	0.38730	0.11775	0.57808	0.17659	0.67790	0.53712	0.67582	0.59893	0.57827	0.85071	1.00000

NO AID

	VAR001	VAR002	VAR003	VAR004	VAR005	VAR006	VAR007	VAR008	VAR009	VAR010	VAR011	VAR012
VAR001	1.00000	-0.11568	0.06903	0.03862	0.03125	0.28910	0.28910	0.28261	0.06894	0.00194	0.01650	0.15917
VAR002	-0.11568	1.00000	-0.56652	0.24791	-0.05249	0.16346	0.09093	-0.27733	0.76733	0.31113	0.74591	0.43366
VAR003	0.06903	-0.56652	1.00000	-0.02119	0.45752	-0.30807	-0.26028	0.29769	-0.65180	0.21800	-0.23093	0.00209
VAR004	0.03862	0.24791	-0.02119	1.00000	-0.13932	0.18620	0.22466	0.58994	0.23856	-0.00025	0.53385	0.56881
VAR005	0.03125	-0.05249	0.45752	-0.13932	1.00000	0.02976	-0.15191	0.14346	-0.09730	0.40020	-0.00470	0.10093
VAR006	0.28910	0.16346	-0.30807	0.18620	0.02976	1.00000	0.79665	0.60106	0.66920	0.23792	0.41674	0.63461
VAR007	0.28910	0.09093	-0.26028	0.22466	-0.15191	0.79665	1.00000	0.50723	0.58600	0.19700	0.34243	0.53282
VAR008	0.28261	-0.27733	0.29769	0.58994	0.14346	0.60106	0.50723	1.00000	0.10349	0.12332	0.23417	0.61909
VAR009	0.06894	0.76733	-0.65180	0.23856	-0.09730	0.66920	0.58600	0.10349	1.00000	0.38930	0.70674	0.61317
VAR010	0.00194	0.31113	0.21800	-0.00025	0.40020	0.23792	0.19700	0.12332	0.38930	1.00000	0.42710	0.56402
VAR011	0.01650	0.74591	-0.23093	0.53385	-0.00470	0.41674	0.34243	0.23417	0.70674	0.42710	1.00000	0.84497
VAR012	0.15917	0.43366	0.00209	0.56881	0.10093	0.63461	0.53282	0.61909	0.61317	0.56402	0.84497	1.00000

NO UNCERTAINTY

	VAR001	VAR002	VAR003	VAR004	VAR005	VAR006	VAR007	VAR008	VAR009	VAR010	VAR011	VAR012
VAR001	1.00000	-0.01595	0.02952	-0.01023	0.03646	0.56760	0.39407	0.43375	0.27940	0.24452	0.20542	0.43402
VAR002	-0.01595	1.00000	-0.62285	0.24903	-0.15030	0.19222	0.09974	-0.24016	0.79552	0.25056	0.70651	0.39301
VAR003	0.02952	-0.62285	1.00000	0.03767	0.51772	-0.20822	-0.17926	0.39699	-0.62963	0.21001	-0.22340	0.11709
VAR004	-0.01023	0.24903	0.03767	1.00000	-0.01630	0.17272	0.19116	0.58677	0.23568	0.10002	0.57360	0.59157
VAR005	0.03646	-0.15030	0.51772	-0.01630	1.00000	0.06776	-0.12570	0.27645	-0.15014	0.37114	-0.07443	0.19255
VAR006	0.56760	0.19222	-0.20822	0.17272	0.06776	1.00000	0.74756	0.60053	0.64394	0.31625	0.49364	0.69305
VAR007	0.39407	0.09974	-0.17926	0.19116	-0.12570	0.74756	1.00000	0.48045	0.53627	0.25347	0.38874	0.55238
VAR008	0.43375	-0.24016	0.39699	0.58677	0.27645	0.60053	0.48045	1.00000	0.08649	0.25061	0.32550	0.68082
VAR009	0.27940	0.79552	-0.62963	0.23568	-0.15014	0.64394	0.53627	0.08649	1.00000	0.40088	0.72761	0.60659
VAR010	0.24452	0.25056	0.21001	0.10002	0.37114	0.31625	0.25347	0.25061	0.40088	1.00000	0.43629	0.58963
VAR011	0.20542	0.70651	-0.22340	0.57360	-0.07443	0.49364	0.38874	0.32550	0.72761	0.43629	1.00000	0.85207
VAR012	0.43402	0.39301	0.11709	0.59157	0.19255	0.69305	0.55238	0.68082	0.60659	0.58963	0.85207	1.00000

OUTCOME

	VAR001	VAR002	VAR003	VAR004	VAR005	VAR006	VAR007	VAR008	VAR009	VAR010	VAR011	VAR012
VAR001	1.00000	-0.03027	0.03734	-0.02035	0.06307	0.57554	0.43572	0.43763	0.20305	0.22299	0.25850	0.42784
VAR002	-0.03027	1.00000	-0.61649	0.22773	-0.12612	0.15794	0.08025	-0.29884	0.77596	0.26721	0.69908	0.37573
VAR003	0.03734	-0.61649	1.00000	0.02999	0.48622	-0.23141	-0.20555	0.38267	-0.65798	0.20844	-0.20800	0.12280
VAR004	-0.02035	0.22773	0.02999	1.00000	-0.07458	0.15584	0.19801	0.57916	0.20697	0.03915	0.53532	0.56409
VAR005	0.06307	-0.12612	0.48622	-0.07458	1.00000	0.01652	-0.15479	0.19844	-0.17152	0.40526	-0.08067	0.15221
VAR006	0.57554	0.15794	-0.23141	0.15584	0.01652	1.00000	0.77871	0.59475	0.65430	0.29902	0.45821	0.66531
VAR007	0.43572	0.08025	-0.20555	0.19801	-0.15479	0.77871	1.00000	0.40827	0.55397	0.22327	0.36480	0.53553
VAR008	0.43763	-0.29884	0.38267	0.57916	0.19844	0.59475	0.40827	1.00000	0.07144	0.20033	0.23448	0.65894
VAR009	0.20305	0.77596	-0.65798	0.20697	-0.17152	0.65430	0.55397	0.07144	1.00000	0.38578	0.70592	0.58727
VAR010	0.22299	0.26721	0.20844	0.03915	0.40326	0.29902	0.22327	0.20033	0.38578	1.00000	0.44913	0.59108
VAR011	0.25850	0.69908	-0.20800	0.53532	-0.08067	0.45821	0.36480	0.28448	0.70592	0.44913	1.00000	0.84922
VAR012	0.42784	0.37573	0.12280	0.56409	0.15221	0.66531	0.53553	0.65894	0.50727	0.59108	0.84922	1.00000

UTILITY

	VAR001	VAR002	VAR003	VAR004	VAR005	VAR006	VAR007	VAR008	VAR009	VAR010	VAR011	VAR012
VAR001	1.00000	-0.10082	0.10991	0.03373	0.02479	0.55245	0.39574	0.50289	0.18636	0.06786	0.20411	0.41987
VAR002	-0.10082	1.00000	-0.60216	0.27849	-0.14491	0.14914	0.06689	-0.26475	0.77692	0.24177	0.70987	0.38555
VAR003	0.10991	-0.60216	1.00000	0.01995	0.51515	-0.23287	-0.23458	0.36749	-0.63852	0.24615	-0.21709	0.11381
VAR004	0.03373	0.27849	0.01995	1.00000	-0.04325	0.14796	0.18661	0.56206	0.24043	0.07405	0.56005	0.57616
VAR005	0.02479	-0.14491	0.51515	-0.04325	1.00000	0.02316	-0.17875	0.22399	-0.18483	0.36947	-0.09654	0.15448
VAR006	0.55245	0.14914	-0.23287	0.14796	0.02316	1.00000	0.76871	0.60870	0.64249	0.24470	0.45492	0.66988
VAR007	0.39574	0.06689	-0.23458	0.18661	-0.17875	0.76871	1.00000	0.48587	0.52898	0.11964	0.35264	0.50738
VAR008	0.50289	-0.26475	0.36749	0.56206	0.22399	0.60870	0.48587	1.00000	0.10038	0.20498	0.29963	0.67513
VAR009	0.18636	0.77692	-0.63852	0.24043	-0.18483	0.64249	0.52898	0.10038	1.00000	0.35546	0.72176	0.60405
VAR010	0.06786	0.24177	0.24615	0.07405	0.36947	0.24470	0.11964	0.20498	0.35546	1.00000	0.41992	0.56551
VAR011	0.20411	0.70987	-0.21709	0.56005	-0.09654	0.45492	0.35264	0.29963	0.72176	0.41992	1.00000	0.84807
VAR012	0.41987	0.38555	0.11381	0.57616	0.15448	0.66988	0.50738	0.67513	0.60405	0.56551	0.84807	1.00000

FULL AID

	VAR001	VAR002	VAR003	VAR004	VAR005	VAR006	VAR007	VAR008	VAR009	VAR010	VAR011	VAR012
VAR001	1.00000	-0.08360	0.09980	0.03673	0.01738	0.60941	0.45224	0.53772	0.24073	0.16679	0.26718	0.47555
VAR002	-0.08360	1.00000	-0.62058	0.27267	-0.17001	0.17474	0.08951	-0.26015	0.78039	0.21964	0.68842	0.36335
VAR003	0.09980	-0.62058	1.00000	0.03935	0.51467	-0.20505	-0.19163	0.39526	-0.64080	0.25006	-0.19282	0.14051
VAR004	0.03673	0.27267	0.03935	1.00000	-0.01135	0.17615	0.20309	0.58331	0.24566	0.09780	0.57855	0.59108
VAR005	0.01738	-0.17001	0.51467	-0.01135	1.00000	0.03215	-0.16262	0.25019	-0.20300	0.37110	-0.09207	0.15967
VAR006	0.60941	0.17474	-0.20505	0.17615	0.03215	1.00000	0.76670	0.60967	0.65391	0.29057	0.49625	0.69455
VAR007	0.45224	0.08951	-0.19163	0.20309	-0.16262	0.76670	1.00000	0.48340	0.54001	0.18298	0.39175	0.54063
VAR008	0.53772	-0.26015	0.39526	0.58331	0.25019	0.60967	0.48340	1.00000	0.09627	0.24940	0.34378	0.70002
VAR009	0.24073	0.78039	-0.64080	0.24566	-0.20300	0.65391	0.54001	0.09627	1.00000	0.34724	0.71543	0.58002
VAR010	0.16679	0.21964	0.25006	0.09780	0.37110	0.29057	0.18298	0.24940	0.34724	1.00000	0.43574	0.58304
VAR011	0.26718	0.68842	-0.19282	0.57855	-0.09207	0.49625	0.39175	0.34378	0.71543	0.43574	1.00000	0.85431
VAR012	0.47555	0.36335	0.14051	0.59108	0.15967	0.69455	0.54063	0.70002	0.58002	0.58304	0.85431	1.00000

DISTRIBUTION LIST

No. of Copies	
5	Director, Engineering Psychology Programs (Code 455) Office of Naval Research 800 North Quincy Street Arlington, VA 22217
12	Defense Documentation Center Cameron Station, Bldg. 5 Alexandria, VA 22314
1	CDR Paul Chatelier Office of the Deputy Under Secretary of Defense OUSDRE (E&LS) Pentagon, Room 3D129 Washington, D.C. 20301
1	CAPT John Duncan Office of the Secretary of Defense (C3I) Pentagon, Room 3C200 Washington, D.C. 20301
1	Dr. Craig Fields Director, Cybernetics Technology Office Defense Advanced Research Projects Agency 1400 Wilson Boulevard Arlington, VA 22209
1	Office of the Chief of Naval Operations OP987H Personnel Logistics Plans Washington, D.C. 20350
1	Dr. A. L. Slafkosky Scientific Advisor Commandant of the Marine Corps Code RD-1 Washington, D.C. 20380
1	Commanding Officer ONR Branch Office ATTN: Dr. J. Lester Bldg. 114, Section D 666 Summer Street Boston, MA 02210
1	Commanding Officer ONR Branch Office ATTN: Dr. Charles Davis 536 South Clark Street Chicago, OL 60605
1	Commanding Officer ONR Branch Office ATTN: Dr. E. Gloye 1030 East Green Street Pasadena, CA 91106

No. of Copies	
1	Analysis and Support Division Code 230 Office of Naval Research 800 North Quincy Street Arlington, VA 22217
2	Naval Analysis Programs Code 431 Office of Naval Research 800 North Quincy Street Arlington, VA 22217
1	Operations Research Program Code 434 Office of Naval Research 800 North Quincy Street Arlington, VA 22217
1	Statistics and Probability Program Code 436 Office of Naval Research 800 North Quincy Street Arlington, VA 22217
1	Information Systems Program Code 437 Office of Naval Research 800 North Quincy Street Arlington, VA 22217
1	Manned Systems Design Code 311 Navy Personnel Research and Development Center San Diego, CA 92152
1	Mr. Mel Moy Code 305 Navy Personnel Research and Development Center San Diego, CA 92152
1	Management Support Department Code 210 Navy Personnel Research and Development Center San Diego, CA 92152
6	Naval Electronics Systems Command Human Factors Engineering Branch Code 4701 Washington, D.C. 20375
6	Director Naval Research Laboratory Technical Information Division Code 2627 Washington, D.C. 20375
1	Mr. Arnold Rubinstein Naval Material Command NAVMAT 08T24 Washington, D.C. 20360

No. of Copies	
1	Dr. John Silva Head, Human Factors Division Naval Ocean Systems Center San Diego, CA 92152
1	Dr. Jesse Orlansky Institute for Defense Analyses 400 Army-Navy Drive Arlington, VA 22202
1	Human Factors Department Code N215 Naval Training Equipment Center Orlando, FL 32813
1	Dr. Gary Poock Operations Research Department Naval Postgraduate School Monterey, CA 92940
1	Dr. Joseph Zeidner, Technical Director U.S. Army Research Institute 500. Eisenhower Avenue Alexandria, VA 22333
1	Dr. Donald A. Topmiller Chief, Systems Effectiveness Branch Human Engineering Division Wright Patterson AFB, OH 45433
1	Dr. H. W. Sinaiko Smithsonian Institution 801 N. Pitt Street Alexandria, VA 2 314
1	Office of the Chief of Naval Operations OP942 Pentagon Washington, D.C. 20350
1	Mr. L. A. Aarons Office of the Chief of Naval Operations OP987C R&D Plans Division Washington, D. 20350
1	Commander Naval Electronic Systems Command C3 Project Office PME 108-1 ATTN: G. Ham on Washington, D.C. 20360
1	CDR P. M. Curran Human Factors Engineering Division Code 604 Naval Air Development Center Warminster, Pa 18974

No. of Copies	
1	Dr. Edgar Johnson Organizations & Systems Research Laboratory U.S. Army Research Laboratory 5001 Eisenhower Avenue Alexandria, VA 22333
1	Dr. David Dianich Chairman, Dept. of Business and Economics Salisbury State College Salisbury, MD 21801
1	Mr. Victor Monteleon Naval Ocean Systems Center Code 230 San Diego, CA 92152
1	Commander, Naval Electronics Systems Command ELEX-03 Washington, D.C. 20360
1	CDR Richard Schlaff NIPSSA Hoffman Bldg. #1 2461 Eisenhower Avenue Alexandria, VA 22331
1	Dr. Chantee Lewis Management Department Naval War College Newport, RI 02840
1	Dr. John Shore Naval Research Laboratory Code 5403 Communications Sciences Division Washington, D.C. 20375
1	Dr. Meredith Crawford American Psychological Association Office of Educational Affairs 1200 17th Street, NW Washington, D.C. 20036
1	Commanding Officer ONR Branch Office ATTN: Mr. R. Lawson 1030 East Green Street Pasadena, CA 91106

No. of Copies	
1	Mr. William Dejka AC 4T Naval Ocean Systems Center San Diego, CA 92152
1	Mr. Merlin Malehorn Office of the Chief of Naval Operations (Op 102) Washington, D.C. 20350
1	Mr. Harold Crane CTEC, Inc. 7777 Leesburg Pike Falls Church, VA 22043
1	Dr. S. D. Epstein Analytics 2500 Maryland Road Willow Grove, PA 19090
1	Dr. Amos Freedy Perceptronics, Inc. 6271 Variel Avenue Woodland Hills, CA 91364
1	Dr. G. Hurst University of Pennsylvania Wharton School Philadelphia, PA 19174
1	Dr. Miley Merkhofer Stanford Research Institute Decision Analysis Group Menlo Park, CA 94025
1	Dr. C. Kelly Decisions and Designs, Inc. 8400 Westpark Drive, Suite 600 P.O. Box 907 McLean, VA 22101
1	Mr. George Pugh Decision Science Applications, Inc. 1401 Wilson Boulevard Arlington, VA 22209
1	Commander, Naval Electronics Systems Command Command and Control Division Code 530 Washington, D.C. 20360

No. of Copies	
1	Dr. Arthur Siegel Applied Psychological Services, Inc. Science Center 404 E. Lancaster Street Wayne, PA 19087
1	Mr. David Walsh Integrated Sciences Corporation 1640 Fifth Street Santa Monica, CA 90401
1	LCDR J. A. Sears Department of MIS College of Business Administration University of Arizona Tucson, AZ 85721
1	I. R. Mirman Asst. for Special Projects HQ AFSC-DL Andrews AFB, MD 20334
1	Mr. Joseph Wohl MITRE Corporation Box 208 Bedford, MA 01730
1	Dr. Kenneth Gardner Applied Psychology Unit Admiralty Marine Technology Establishment Teddington, Middlesex TW11 OLN ENGLAND
1	Mr. Tim Gilbert The MITRE Corporation 1820 Dolly Madison Blvd. McLean, VA 22102
1	Mr. Leslie Innes Defense & Civil Institute of Environmental Medicine P.O. Box 2000 Downsview, Ontario M3M 3B9 CANADA
1	Mr. M. L. Metersky Naval Air Development Center Code 5424 Warminster, PA 18974